



Department of Performance Monitoring and Evaluation

Report on the Assessment of Government Evaluations

| | |
|-----------------------------|---|
| Evaluation Title: | Developing and Evaluating the First Phase of the Grade 12 Mind the Gap Study Guide Series |
| Evaluation Number: | 420 |
| Evaluation Completion Date: | 21 October 2013 |
| Period of Evaluation: | September 2012 to October 2013 |
| Submitted: | 30 January 2014 by Ray Basson |
| Approved: | 03 February 2014 by Mike Leslie |

Evaluation Details

| | |
|-----------------------------|---|
| Evaluation Title: | Developing and Evaluating the First Phase of the Grade 12 Mind the Gap Study Guide Series |
| Evaluation Number: | 420 |
| Evaluation Completion Date: | 21 October 2013 |
| Created: | 29 November 2013 by Mike Leslie |
| Submitted: | 30 January 2014 by Ray Basson |
| Approved: | 03 February 2014 by Mike Leslie |
| Period of Evaluation: | September 2012 to October 2013 |
| Known Cost: | R 0.00 |
| Known Cost Type: | Estimate |
| Initiated By: | DBE materials developer |
| Initiated By Internal: | Yes |
| Undertaken By: | Dr S Taylor, DBE |
| Undertaken By Internal: | Yes |

Assessors

| | |
|------------|----------------------------|
| Ray Basson | raymond.b.basson@gmail.com |
|------------|----------------------------|

Assessment Documents

| Document Name: | Document Type: | Added By: | Added On: |
|--|---|-------------|------------------|
| Mind the Gap Technical Report 2013.pdf | Evaluation report | Mike Leslie | 29 November 2013 |
| PDG 21014 Mind the Gap Technical Report 2013.txt | Evaluation report | Ray Basson | 23 January 2014 |
| Assessment Report at Moderation.pdf | Assessment Report at Moderation | Mike Leslie | 28 January 2014 |
| PDG 2014 MtG Concept document.txt | Terms of Reference (ToR) for the evaluation | Ray Basson | 29 January 2014 |

Quality Assessment Summary

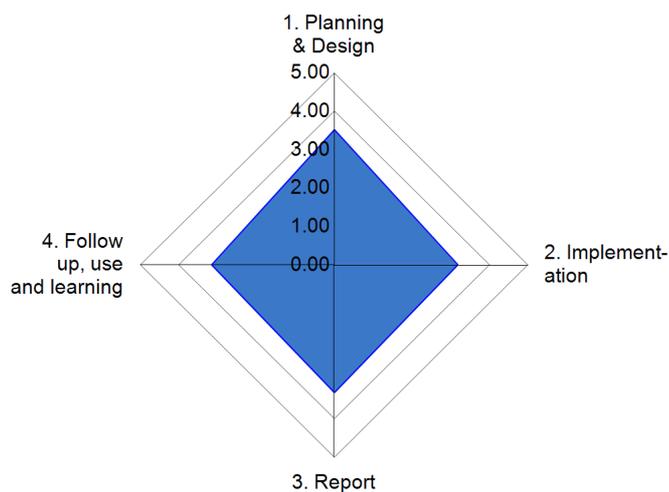
The aim of the Mind the Gap impact evaluation was to establish broadly if its learner self-instruction materials made a difference to Grade 12 learner performance in the high stakes NSC examination, and consider taking them to scale if they did. It asked: did the materials improve learner performance?, was the improvement had in all 4 subjects?, and, which learners benefited most from the intervention? Adopting an impact/RCT design and using secondary data sourced from a website with NSC results, evaluators found the materials did improve performance, that performance was improved in 2 of the 4 subjects, and that more able learners benefited most from using the materials. Findings were not statistically significant, and thus need to be treated cautiously especially in discussion to take the materials to scale. In this regard authors should also have noted the usual practice in RCTs to produce statistically significant findings and test for reliability, amongst others, before a discussion of use. As it stands, the evaluation appears a rigorous application of an impact evaluation which has been carefully implemented, but its claims may overstate what can be done with its findings. As impact evaluation is a recognized approach by government and in the evaluation literature, it has the advantage of being open to debate, unlike approaches which are not. And it should be noted this application of impact evaluation is unusual as it relies on secondary sources of data and not, as in more usual applications, on test development, testing on site, and the like. The assessment score of 3.30 appears about right for the evaluation as it stands, which is likely to have benefited from critical comment from peers external to the DBE.

Quality Assessment Scores

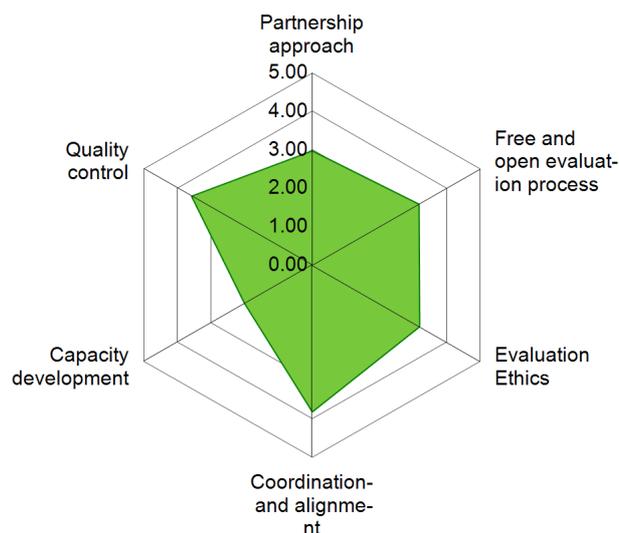
| Phase of Evaluation | Score |
|--------------------------------|-------------|
| 1. Planning & Design | 3.52 |
| 2. Implementation | 3.19 |
| 3. Report | 3.31 |
| 4. Follow-up, use and learning | 3.15 |
| Total | 3.30 |

| Overarching Consideration | Score |
|----------------------------------|-------------|
| Partnership approach | 2.97 |
| Free and open evaluation process | 3.18 |
| Evaluation Ethics | 3.20 |
| Coordination and alignment | 3.83 |
| Capacity development | 2.00 |
| Quality control | 3.58 |
| Total | 3.30 |

Scores: Phases of Evaluation



Scores: Overarching Considerations



| Phase of Evaluation | Area of Evaluation | Score |
|--------------------------------|--|-------------|
| 1. Planning & Design | 1.1. Quality of the TOR | 3.50 |
| 1. Planning & Design | 1.2. Adequacy of resourcing | 3.82 |
| 1. Planning & Design | 1.3. Alignment to policy context and background literature | 4.40 |
| 1. Planning & Design | 1.4. Appropriateness of the evaluation design and methodology | 3.22 |
| 1. Planning & Design | 1.5. Project management (Planning phase) | 3.00 |
| 2. Implementation | 2.1. Evaluation ethics and independence | 2.50 |
| 2. Implementation | 2.2. Participation and M&E skills development | 2.63 |
| 2. Implementation | 2.3. Methodological integrity | 3.80 |
| 2. Implementation | 2.4. Project management (Implementation phase) | 4.00 |
| 3. Report | 3.1. Completeness of reporting structure | 3.17 |
| 3. Report | 3.2. Accessibility of content | 3.47 |
| 3. Report | 3.3. Robustness of findings | 3.32 |
| 3. Report | 3.4. Strength of conclusions | 3.07 |
| 3. Report | 3.5. Suitability of recommendations | 3.54 |
| 3. Report | 3.6. Consideration of reporting risks and ethical implications | 3.70 |
| 3. Report | 3.7. Project management (Reporting phase) | 2.00 |
| 4. Follow-up, use and learning | 4.1. Resource utilisation | 3.40 |
| 4. Follow-up, use and learning | 4.2. Evaluation use | 3.10 |
| Total | Total | 3.30 |

1. Planning & Design

1.1. Quality of the TOR

| | |
|-----------------------|--|
| Standard: | 1.1.1. The evaluation was guided by a well-structured and complete TOR or a well-structured and complete internal evaluation proposal |
| Comment and Analysis: | The evaluation was guided by an internal concept document, which outlined the impact evaluation design, methods and analysis procedures. Being internal, it was not as formal or structured as in a TOR for an external service provider. |
| Rating: | 3 |
| Standard: | 1.1.2. The purpose of the evaluation stated in the TOR (or an internal evaluation proposal) was clear and explicit |
| Comment and Analysis: | The purpose of the evaluation was clear, to evaluate the effects of the materials on learner performance with the view to scaling-up the MtG intervention in 2014. |
| Rating: | 4 |
| Standard: | 1.1.3. The evaluation questions in the TOR (or an internal evaluation proposal) were clearly stated and appropriate for addressing the evaluation purpose |
| Comment and Analysis: | The evaluation questions were appropriate to the evaluation purpose as questions asked if the materials made a difference to learner performance, if difference was the same across all subjects and who benefited from the materials, with the view to providing data for a decision to/not to scale-up the intervention in 2014. |
| Rating: | 4 |
| Standard: | 1.1.4. 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: | A RCT or impact evaluation approach is widely accepted in the evaluation community to have strengths over other approaches for establishing the effects of an intervention on subjects. Adopting the approach, thus, was well suited to the purpose and scope of the evaluation. |
| Rating: | 4 |
| Standard: | 1.1.5. The TOR (or an internal evaluation proposal) identified the intended users of the evaluation and their information needs |
| Comment and Analysis: | The intended users of the findings of the evaluation were the DBE materials developers and colleagues in the DGs office responsible for the intervention and its use in 2014. They wanted to know what effects the materials had on learner performance in the NSC examination. |
| Rating: | 3 |

| | |
|-----------------------|--|
| Standard: | 1.1.6. Key stakeholders were involved in the scoping of the TOR and choosing the purpose of the evaluation |
| Comment and Analysis: | Key stakeholders were internal to the DBE. The materials developers specifically were involved constantly with the evaluator in a collegial, rather than formal, process for choosing the purpose of and scoping the evaluation. |
| Rating: | 3 |

1.2. Adequacy of resourcing

| | |
|-----------------------|--|
| Standard: | 1.2.1. The evaluation was adequately resourced in terms of time allocated |
| Comment and Analysis: | The evaluation had sufficient time to complete its work. This was determined mainly by the rate at which materials were produced, the availability of the NSC results, and the evaluators' other in-house DBE assignments. |
| Rating: | 4 |

| | |
|-----------------------|---|
| Standard: | 1.2.2. The evaluation was adequately resourced in terms of original budget |
| Comment and Analysis: | As the data for the evaluation was obtained from the NSC database and materials development had its own budget [R75,000-00], evaluation costs were minimal and confined to expenses, such as salaries, paid by the DBE. None of the usual RCT procedures and costs including test development, testing learners, training to use materials, visits to monitor progress by officials, hiring fieldworkers, applied in this case which usually make impact evaluation expensive. However, the evaluator felt that additional funding for some fieldwork on how the materials were used, would have been useful. |
| Rating: | 4 |

| | |
|-----------------------|---|
| Standard: | 1.2.3. The evaluation was adequately resourced in terms of staffing and skills sets |
| Comment and Analysis: | As employee of the DBE with responsibility for evaluation, the evaluator alone largely comprised the evaluation team and had the skills set required to do the analysis of secondary data to establish effects of materials on learner performance. |
| Rating: | 4 |

| | |
|-----------------------|--|
| Standard: | 1.2.4. Where appropriate, the evaluation planned to incorporate an element of capacity building of partners/staff responsible for the evaluand |
| Comment and Analysis: | Capacity building of colleagues linked to the evaluation was carried out as ordinarily done in the DBE, but is not specified in the concept document for the evaluation. This evaluation, however, was used as a case in the ongoing evaluation workshops conducted by the evaluator in the normal course of his duties. |
| Rating: | 2 |

1.3. Alignment to policy context and background literature

Standard: 1.3.1. There was evidence that a review of the relevant policy and programme environments had been conducted and used in planning the research

Comment and Analysis: Literature reviews were conducted and reports studied of relevant programme and policy environments informing this research. The Diagnostic Report on Learner Performance in 2011 (DBE) was reviewed for an analysis of learners examination responses and the essential content required, to develop self-instruction learner materials in 4 subjects targeting underperforming learners, to improve their performance in the high stakes NSC examination. In contrast with expert developed learner materials researched and reported in the literature, Mind the Gap produced ordinary innovation materials which involved teachers from low performing schools as experts on underperforming schools, to design learner self-instruction materials requiring no support from teachers. CAPS policy documents were briefly reviewed to align most recent policy developments with learner self-instruction materials. Part 1 of the study draws on relevant theorists including Wittgenstein, Freire, Bordieu and others, to make the case conceptualizing the design of the learner self-instruction materials, to improve grades. This is a strength of this evaluation.

Rating: 5

Standard: 1.3.2. There was evidence of a review of appropriate literature having been conducted and used in planning the research

Comment and Analysis: The literature on impact evaluation, or RCTs, in Africa and America provides a not too optimistic background for materials improving learner performance in this evaluation. The review nevertheless provides a background for planning this research and an opportunity to further this debate with respect to learner self-instruction materials, by establishing effects, and more specifically, if these support or run counter to findings in the review and improve instruction.

Rating: 4

1.4. Appropriateness of the evaluation design and methodology

Standard: 1.4.1. 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: There was explicit reference to the theory of change of the evaluand in planning the evaluation. The wording in the theory of change, however, presumes change will happen and be positive. It would be more realistic and persuasive if steps in the theory of change persuaded to the possibility of change resulting from learners' use of the materials.

Rating: 3

Standard: 1.4.2. Key stakeholders were consulted on the design and methodology of the evaluation

Comment and Analysis: Stakeholders internal to the DBE, largely from the DGs office, were consulted collegially, rather than formally, on the design and methods of the evaluation.

Rating: 2

Standard: 1.4.3. The planned methodology was appropriate to the questions being asked

Comment and Analysis: Using a RCT methodology to gauge if these learner self-instruction materials improve grades of underperforming learners in the NSC in Geography, Life Sciences, Economics, Accounting, is wholly appropriate to the questions asked: what is their impact?, is impact the same across all subjects?, which learners benefit most from these materials?

Rating: 4

Standard: 1.4.4. Sampling was appropriate and adequate given focus and purpose of evaluation

Comment and Analysis: The use of a computerized lottery sampling procedure was appropriate to the purpose of the evaluation and ensured schools were randomly selected and allocated to both the sample and control groups. It should be noted, however, that this sampling procedure making it a RCT, differs from more usual RCT sampling where comparability is established first, and is followed by control of variables, instrument validation, testing, testing for statistical significance and for reliability, and the like. These procedures are usually followed before making claims about effects and their use. In the Mind the Gap case, secondary data on NSC results are utilised to establish difference, are tested for heterogeneity effects, spurious effects, robustness of the sample, and the like. Statistically significant results were not attained and a discussion of these materials going to scale moved more quickly than would have been the case in a standard RCT with testing. Clearly cautions need to be put in place, despite the randomizing sampling procedure.

Rating: 4

Standard: 1.4.5. There was a planned process for using the findings of the evaluation

Comment and Analysis: With the evaluation being planned collegially, using the findings coincided with Department of Basic Education [DBE] processes for scaling-up the MtG intervention. With the evaluation providing an estimate of additional numbers of NSC passes had these materials had been made available earlier, and numbering around 5000, added impetus to the DBE's process for scaling up the MTG programme.

Rating: 3

1.5. Project management (Planning phase)

Standard: 1.5.1. The inception phase was used to develop a common agreement on how the evaluation would be implemented

Comment and Analysis: Whilst there was no formal inception phase in the evaluation, informal collegial exchanges between evaluator and materials developer in the DBE from its inception resulted in common agreements on how the evaluation would be implemented, and were continued throughout.

Rating: 3

2. Implementation

2.1. Evaluation ethics and independence

Standard: 2.1.1. Where data was gathered in contexts where ethical sensitivity is high, appropriate clearance was achieved through an ethics review board; e.g. in evaluation involving minors, institutions where access usually requires ethical or bureaucratic clearance, and situations where assurances of confidentiality was offered to participants

Comment and Analysis: As the MtG intervention was an official project in the DBE and was already sending its materials to underperforming schools in the country as well as uploading them on the web, bureaucratic clearance was assumed by the evaluator. No formal permission was sought, and possibly not required. As there was no testing of learners/individuals, ethical clearance was not required.

Rating: 3

Standard: 2.1.2. Where external, the evaluation team was able to work freely without significant interference

Comment and Analysis: This was an internal evaluation.

Rating: N/A

Standard: 2.1.3. The evaluation team was impartial and there was no evidence of conflict of interest

Comment and Analysis: As internal evaluation, the evaluation runs the risk of being partial and serving the interests of the DBE. The evaluator argues it was impartial and with no conflict of interest despite working closely with the MtG materials developer. However, words are used in the report which presume what the evaluation is designed to establish early in the report and prior to a discussion of the evidence. This is surprising for a study premised on objectivity, and when encountered raises questions about its impartiality.

Rating: 2

2.2. Participation and M&E skills development

Standard: 2.2.1. Key stakeholders were consulted through a formalised mechanism or institutional arrangement

Comment and Analysis: Key stakeholders in the DBE were consulted, collegially however, rather than through a formal institutional arrangement.

Rating: 3

Standard: 2.2.2. Where appropriate, an element of capacity building of partners responsible for the evaluand was incorporated into the evaluation process

Comment and Analysis: No formal capacity building of partners responsible for the evaluand was built into the evaluation process. But DBE parties in the evaluation received some on-site inputs in evaluation, and the MtG RCT was cited as an evaluation case in training the evaluator does internally for DBE staff.

Rating: 2

Standard: 2.2.3. Where appropriate, the evaluation team incorporated an element of skills development amongst the evaluators (e.g. students, interns, recent graduates, etc)

Comment and Analysis: As the evaluator is a measurement specialist and there were no interns and recent graduates on the team, no skills development of evaluators was needed or incorporated into the evaluation.

Rating: N/A

Standard: 2.2.4. Peer review of the agreed evaluation design and methodology occurred prior to undertaking data collection

Comment and Analysis: The design and methods of the evaluation was peer reviewed by colleagues internal to the DBE. Before implementation, the design and methods were presented to DBE management and staff as well as at a conference from which feedback was received and taken into account.

Rating: 3

2.3. Methodological integrity

Standard: 2.3.1. The methods employed in the process of the evaluation were consistent with those planned

Comment and Analysis: As secondary data sources were planned to be used, these were followed unproblematically in executing the evaluation.

Rating: 4

Standard: 2.3.2. A pilot of data collection instrumentation occurred prior to undertaking data collection

Comment and Analysis: Piloting instruments/tests was not part of the evaluation, and thus not part of the process.

Rating: N/A

Standard: 2.3.3. Data collection was not compromised by fieldwork-level problems or unplanned diversions from original intentions

Comment and Analysis: As secondary data were used, no fieldwork was part of the study.

Rating: N/A

Standard: 2.3.4. Forms of data gathering were appropriate given the scope of evaluation
Comment and Analysis: The use of data on the NSC examinations 2010-12 was appropriate for establishing if the pass rate in the treatment schools increased over the pass rate in control schools.
Rating: 4

Standard: 2.3.5. The data analysis approach and methods were appropriate and sufficient given the purposes of the evaluation
Comment and Analysis: The use of regression analysis techniques on data obtained from the NSC 2011-12 results to establish the effect of materials on learners, appears appropriate. Data analysis followed confirmation of delivery of materials to all control schools and the materials being uploaded on the DBE website for wider access and use. The latter, the authors acknowledge, may have introduced a slight possibility of a contamination effect. The finding of improvement in 2 subjects, Geography and Life Sciences, was tested for heterogenous effects, such as gender, and multiple placebo tests were run which provides assurance that impacts were not the result of spurious data gathering effects in these subjects.
Rating: 4

Standard: 2.3.6. Key stakeholders were significantly engaged as part of the methodology
Comment and Analysis: Key stakeholders internal to the DBE were engaged collegially, but not significantly, as part of the methodology.
Rating: 3

Standard: 2.3.7. The methodology included engaging beneficiaries appropriately as a key source of data and information
Comment and Analysis: As the evaluation was conceived and carried out remotely and secondary data sources were used, beneficiaries were not engaged as a key source of data.
Rating: N/A

2.4. Project management (Implementation phase)

Standard: 2.4.1. The evaluation was conducted without significant shifts to scheduled project milestones and timeframes
Comment and Analysis: There were no significant shifts in project milestones and timeframes.
Rating: 4

3. Report

3.1. Completeness of reporting structure

Standard: 3.1.1. Executive summary captures key components of the report appropriately

Comment and Analysis: The Executive Summary reports the process of materials development, conceptualization and design as ordinary innovation in education, and thereafter the evaluation. As outlined below, the Summary captures key features in the evaluation, design through conclusions, but it does not mention that results were not statistically significant, and it lacks succinctness and dispassionate reporting of the study. As a result, the discussion of scaling-up the intervention appears premature. Summarizing, the RCT tested the impact of the materials of 4 subjects written in English: Geography, Life Sciences, Accounting and Economics. To demonstrate impact, the design of the RCT required a randomized treatment and control group and an outcome measure, such as NSC results. Three evaluation questions ask: can the learning materials improve performance in the high stakes NSC examination?, is impact the same across subjects?, and, who benefits most from these materials? The findings indicate impact on performance in 2 of the 4 subjects, Geography and Life sciences, but not in Economics and Accounting. Impact was not the same across subjects, and lower amongst some seriously underperforming schools, in Geography mainly. Increased performance in 2 subjects would lead to an increased probability of attaining an NSC certificate. It concludes that this innovation can be expected to significantly impact performance going forward, that increased performance in 2 subjects provides solid evidence these materials can accomplish its intended purpose, and provides a clear policy warrant to expand the innovation to scale in 2014 as well as develop materials in 4 more subjects aligned with the CAPS statements.

Rating: 3

Standard: 3.1.2. The context of the development intervention is explicit and presented as relevant to the evaluation

Comment and Analysis: The Mind the Gap materials were developed in 4 subjects to address widespread learner underperformance in the NSC examination in the country. The context is relevant to the evaluation as it aims to establish if these self-instruction learner materials improve grades in the NSC examination, and if they do, to estimate how many learners may have passed the examination had these materials been available earlier, the intention being to take materials to scale. The wider context of RCT evaluations of materials based interventions in Africa and elsewhere presented in the evaluation, however, is not too optimistic that materials make a difference to learner performance.

Rating: 4

Standard: 3.1.3. There is a clear rationale for the evaluation questions

Comment and Analysis: The rationale for the evaluation questions derives from a primary interest to establish if MtG materials make a difference to learner performance, and if they improve performance, to use this as evidence for decisions to take the intervention to scale.

Rating: 3

| | |
|-----------------------|--|
| Standard: | 3.1.4. The scope or focus of the evaluation is apparent in the report |
| Comment and Analysis: | The focus of the evaluation is to establish the impact of 4 sets of learning materials developed in the Mind the Gap intervention on Grade 12 learners in randomly selected treatment and control schools in Mpumalanga. Whilst the study checked materials were delivered to all treatment school, outside its scope was a check to see that the materials were handed to learners by teachers and checks that learners actually used these materials during the one to two month window between delivery and writing an examination. |
| Rating: | 4 |
| Standard: | 3.1.5. A detailed methodology is outlined in the relevant section of the report to the point that a reader can understand the data collection, analysis and interpretation approaches used |
| Comment and Analysis: | Methods are detailed and appropriate in this study, and help understand the data, its analysis and interpretation. It is clear from the report that secondary data sources/NSC results is used, so no detailed data gathering methods are outlined in the report. Data analysis procedures are discussed in detail including, regression analyses, establishing heterogenous effects, placebo treatments, and the like to strengthen findings and estimates, as well as establish if any spurious effects were found. Less detailed discussion is given for establishing the level of statistical significance in the primary finding, unit cost of the materials, and the number of additional passes there may have been in 2011 assuming improved performance in two subjects. |
| Rating: | 4 |
| Standard: | 3.1.6. Acknowledgement of limitations of all aspects of the methodology and findings are clearly and succinctly articulated |
| Comment and Analysis: | The study explicitly reminds of limitations inhering in qualitative evaluations, as these in the DBE's experience, have not provided definitive data measuring impact, thus providing the reason for using a RCT in this evaluation. However, no limitations of RCTs appear acknowledged, the obvious being the difficulty of establishing statistically significant effects, as in this case, difficulty in selecting randomly comparable treatment and control groups, controlling for variables, and the like. While absent here, a discussion of statistical significance is pertinent for gauging the strength of findings indicating positive effect of materials on performance as well as for scaling-up the intervention. Also, in places the language talks up the evaluation, as in 2.5 Conclusions where this RCT is referred to as "good" impact evaluation, highlighting the need of critical comment about the approach and its procedures. |
| Rating: | 2 |
| Standard: | 3.1.7. Key findings are presented in a clear way; they are made distinct from uncertain or speculative findings; and unused data is not presented in the body of the report |
| Comment and Analysis: | Key findings are presented clearly in the appropriate section of the study and its conclusions. Where findings are presented in response to questions, findings are well grounded in data and persuasive. Where discussion does not appear to respond to a question, such as the discussion of RCT trials as approach to use in government in the conclusion, assertions appear to be made without explicit prior discussion and evidence and, thus, appear speculative. And, the study appears to have presented all data extracted from the database on NSC results. |
| Rating: | 3 |

Standard: 3.1.8. Conclusions and recommendations are clear and succinctly articulated

Comment and Analysis: Conclusions would have been strengthened had these been closely linked to the evaluation questions. This would also have resulted in fuller discussion of issues raised in questions, and to avoid much discussion on issues related to RCTs in government, their cost and the like. Recommendations do not appear in a formal section in the report, are made and clear, but with little discussion.

Rating: 2

3.2. Accessibility of content

Standard: 3.2.1. The final evaluation report is user-friendly, written in accessible language and its content follows a clear logic

Comment and Analysis: The final report is user-friendly, written in accessible language and is structured by the logic entailed in the evaluation questions. Conceptualizing the materials development process in Part 1, is longer than would usually be expected in an evaluation report on a RCT.

Rating: 3

Standard: 3.2.2. Quality of writing and presentation is adequate for publication including: adequate layout and consistent formatting; complete sentences and no widespread grammatical or typographical errors; consistency of style and writing conventions (e.g. tense, perspective (first person, third person); levels of formality; references complete and consistent with cited references in reference list and vice versa; etc.

Comment and Analysis: The quality of writing and presentation is adequate. Layout and formatting is consistent, and there don't appear to be widespread grammatical and typographical errors. Referencing in the body of the text and in the list of References appear consistent and complete.

Rating: 3

Standard: 3.2.3. 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.)

Comment and Analysis: Appropriate conventions are used in the presentation of data. Values, statistically insignificant findings and the like are reported, and the language of write-up appears appropriate to reporting statistics.

Rating: 4

Standard: 3.2.4. The use of figures and tables is such that it supports communication and comprehension of results; and data reported in figures and tables are readily discernible and useful to a reader familiar with data presentation conventions

Comment and Analysis: Many tables and figures presented in the report support communication and comprehension of results. They clearly present the data, which is readily discernible and useful to a reader familiar with statistical data presentation procedures.

Rating: 4

3.3. Robustness of findings

Standard: 3.3.1. Data analysis appears to have been well executed
Comment and Analysis: Data analysis appears to have been explained and well executed and is accessible to the reader.
Rating: 4

Standard: 3.3.2. Findings are supported by available evidence
Comment and Analysis: Findings are supported by available evidence and checks made on the evidence.
Rating: 4

Standard: 3.3.3. The evidence gathered is sufficiently and appropriately analysed to support the argument
Comment and Analysis: The evidence gathered from the database on NSC results appears to be sufficient and appropriately analysed to support the argument namely, qualified improvement following the use of MtG materials.
Rating: 4

Standard: 3.3.4. There is appropriate recognition of the possibility of alternative interpretations
Comment and Analysis: Little recognition of the possibility of alternative interpretations appears in the discussion of findings, and none appear in the report.
Rating: 2

Standard: 3.3.5. The report appears free of significant methodological and analytic flaws
Comment and Analysis: The report appears to be free of significant methodological and analytic flaws.
Rating: 4

Standard: 3.3.6. Relevant limitations of the evaluation are noted
Comment and Analysis: Few if any relevant limitations are noted in the report.
Rating: 1

3.4. Strength of conclusions

Standard: 3.4.1. Conclusions are derived from evidence

Comment and Analysis: Conclusions are derived from evidence in the form of secondary data derived from a database of NSC results. A comparison of results from the treatment and control groups established the materials improved the performance of underperforming learners in 2 subjects in the 2012 NSC examination. This is a strength of impact evaluation. But, it needs to be qualified by comments on the strength of these findings--they were not statistically significant--and the implications of this for the intervention going forward and especially taking it to scale.

Rating: 4

Standard: 3.4.2. Conclusions take into account relevant empirical and/or analytic work from related research studies and evaluations

Comment and Analysis: Conclusions take into account other relevant RCT evaluations on the effect of learning materials on learner performance in Africa and elsewhere. As wider context, relevant and related research provides a not too optimistic backdrop on materials improving performance. Thus, the finding of improvement in 2 subjects in this impact evaluation is both pertinent and interesting, but also needs further testing because it differs from a trend established in related studies.

Rating: 4

Standard: 3.4.3. Conclusions address the original evaluation purpose and questions

Comment and Analysis: The conclusions address the original purpose and questions of the evaluation. The evaluation questions, however, don't appear to explicitly structure conclusions to keep them tight and persuasive as intended in a RCT.

Rating: 3

Standard: 3.4.4. Conclusions are drawn with explicit reference to the intervention logic or theory of change

Comment and Analysis: No explicit reference to the intervention logic is made in the discussion of findings in the report.

Rating: 1

3.5. Suitability of recommendations

Standard: 3.5.1. Recommendations are made in consultation with appropriate sectoral partners or experts

Comment and Analysis: Recommendations are made in the body of the evaluation report, but not in a section for that purpose. The main recommendation for a policy warrant to scale up the intervention was made internally, not with sectoral partners external to the DBE. The evaluator notes, however, that this recommendation may have gained strength in the DBE in the absence of evaluations of other completing interventions.

Rating: 3

| | |
|-----------------------|--|
| Standard: | 3.5.2. Recommendations are shaped following discussions with relevant government officials and other relevant stakeholders |
| Comment and Analysis: | Recommendations are shaped collegially with officials internal to the DBE. |
| Rating: | 3 |
| Standard: | 3.5.3. Recommendations are relevant to the current policy context |
| Comment and Analysis: | Recommendations are relevant to the current policy context, where there is a renewed emphasis on subject content and debate about the pass rate in the NSC examinations. |
| Rating: | 4 |
| Standard: | 3.5.4. Recommendations are targetted at a specific audience sufficiently - are specific, feasible, affordable and acceptable |
| Comment and Analysis: | Recommendations target learner performance in the Grade 12 NSC examination. Recommendations were not intended to be specific and lead to further materials development, but were general for the MtG subject materials to be more widely used. In respect to MTG intervention in general, scaling-up the intervention appears feasible, affordable and acceptable. |
| Rating: | 4 |

3.6. Consideration of reporting risks and ethical implications

| | |
|-----------------------|---|
| Standard: | 3.6.1. Peer review of the draft evaluation report occurred prior to finalisation of the evaluation report |
| Comment and Analysis: | No formal peer review by experts external to the DBE occurred prior to finalization of the evaluation report. Internal collegial review did occur. |
| Rating: | 3 |
| Standard: | 3.6.2. 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: | As secondary data is used in the evaluation, informed consent was not required or obtained to ensure confidentiality. |
| Rating: | N/A |
| Standard: | 3.6.3. There are no risks to participants in disseminating the original report on a public website |
| Comment and Analysis: | As names of individuals who participated in the study are not referred to or mentioned in the report, there appear to be little or no risk to participants in disseminating the original report on a public website. |
| Rating: | 4 |

| | |
|-----------------------|--|
| Standard: | 3.6.4. There are no unfair risks to institutions in disseminating the original report on a public website |
| Comment and Analysis: | As names of participating institutions are not referred to or mentioned, no unfair risks to institutions seem likely in disseminating the original report on a public website. |
| Rating: | 4 |

3.7. Project management (Reporting phase)

| | |
|-----------------------|---|
| Standard: | 3.7.1. A project closure meeting that reflected on the challenges and strengths of the evaluation process occurred |
| Comment and Analysis: | No formal meeting was organized to reflect on the evaluation, its strengths and weaknesses. The evaluation report, however, went live with other reports on DBE interventions for ongoing comment and debate. |
| Rating: | 2 |

4. Follow-up, use and learning

4.1. Resource utilisation

Standard: 4.1.1. The evaluation was completed within the planned timeframes

Comment and Analysis: The evaluation was completed within the planned timeframe, and was completed whilst fulfilling other DBE assignments and routine tasks which crossed the evaluators desk. Despite there being a waiting period for NSC results to be uploaded onto the website when the evaluator was able to turn his mind to routine activities, the timeline appears to be more than adequate for a RCT using secondary sources.

Rating: 3

Standard: 4.1.2. The evaluation was completed within the agreed budget

Comment and Analysis: The evaluation was completed within the budget. No additional costs were incurred.

Rating: 4

4.2. Evaluation use

Standard: 4.2.1. Results of the evaluation have been presented to all relevant stakeholders

Comment and Analysis: The results of the evaluation have been widely presented to a wide range of stakeholders within the DBE, to the M&E unit, Curriculum Planning unit, management and others.

Rating: 4

Standard: 4.2.2. A reflective process has been undertaken by the steering committee (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 specific reflective process has been undertaken in the DBE to strengthen future evaluations. Collegial communications and debate reflect all the time on such issues, which in effect is how reflection occurred in the evaluation.

Rating: 2

Standard: 4.2.3. 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: In the interview, the evaluator felt it had added symbolic value to the materials development programme, as it lent credibility to it and the materials it develops just by virtue of it having been evaluated. The materials developer said the evaluation added weight to debates about the materials and about building evaluation in to new projects in the DBE.

Rating: 4

| | |
|-----------------------|---|
| Standard: | 4.2.4. The evaluation study is of conceptual value in understanding what has happened and possibly in shaping policy and practice |
| Comment and Analysis: | In the interview, the evaluator indicated the evaluation had conceptual value for shaping policy as it assisted understanding the effects of materials on performance. It also helped promote a culture of evaluation in the DBE. |
| Rating: | 3 |
| Standard: | 4.2.5. Development of a draft improvement plan has been started, but not completed, based on the findings and recommendations set out in the evaluation |
| Comment and Analysis: | The development of a draft improvement plan doesn't appear to have started. But the finding of the evaluation have shown the DBE that not all interventions make a difference, or are of the same quality, and the like. |
| Rating: | 2 |
| Standard: | 4.2.6. The report is publicly available (website or otherwise published document), except where there were legitimate security concerns *Note: only apply if sufficient time has elapsed since completion of the evaluation |
| Comment and Analysis: | The report is publicly available on the DBE website |
| Rating: | 4 |
| Standard: | 4.2.7. There is clear evidence of instrumental use - that the recommendations of the evaluation were implemented to a significant extent *Note: only apply if sufficient time has elapsed since completion of the evaluation |
| Comment and Analysis: | The evaluator was not sure that recommendations were implemented or that MtG will be scaled-up. However, this process is in motion with due influence by the findings. |
| Rating: | 3 |
| Standard: | 4.2.8. There is clear evidence that the evaluation has had a positive influence on the evaluand, its stakeholders and beneficiaries over the medium to long term *Note: only apply if sufficient time has elapsed since completion of the evaluation |
| Comment and Analysis: | The evaluator was not sure about positive influence on the evaluand, stakeholders and beneficiaries. However, the materials developer made it clear that findings were used shortly after being made known within the DBE, to revise particularly Accounting and Economics materials. The former was found to be short of content and thus that coverage of the syllabus had to be rectified. In the latter, a cognitive plateau was found to lock-out learners at a point when learners became saturated with the volume of substantive content to learn. She made the point, too, that it has had the effect of evaluations gaining traction in debates within the DBE, and of the importance of building evaluation into a programme from inception. |
| Rating: | 3 |

References

Design of the Impact Evaluation of the "Mind the Gap" Study Guides.

Technical Report - Developing and Evaluating the First Phase of the Grade 1 Mind the Gap Study Guide Series.

List of Interviewees

Dr S Taylor, DBE, telephonic interview, 22 January 2014.

Dr P Watson, DBE, MtG materials developer, telephonic interview, 30 January 2014.