



Class ranking of secondary schools in the North West province of South Africa

G Kent*

HA Kruger[†]

JV du Toit*

Received: 16 September 2016; Revised: 14 November 2016; Accepted: 21 November 2016

Abstract

The education system in South Africa finds itself in a dismal position. Existing research reports and general news bulletins suggest that schools in South Africa suffer from a general lack of efficiency measures and therefore underperform in comparison with those in other countries. This paper proposes a class-ranking technique as a means to evaluate schools. The technique implements Pareto optimality principles and utilises a mathematical model that is based on output variables (*e.g.* pass rates) only. The model also provides for the construction of a step-by-step improvement plan for underperforming schools. The suggested framework was applied to 54 secondary schools in one of the four major municipal districts in the North West province of South Africa. Results are contrasted with the application of an ordinary DEA model that is based on input and output variables and showed that the suggested approach offers a more reliable alternative to evaluate, monitor and improve the performance of schools ranked as inefficient.

Key words: Class ranking, DEA, Pareto optimality, school efficiency.

1 Introduction

South Africa is a multiracial country with a new and young democracy of just over 21 years old. The country is facing a number of national problems in various socio-economic areas. One of the most prominent problems, appearing to be an on-going concern for both the national government and a significantly large proportion of South African citizens, is the poor performance of schools. Spaull [30], for example, conducted a study into the quality of education in South Africa during 1994–2011 and concluded that “the South African education system is grossly inefficient, severely underperforming and egregiously unfair”.

*School of Computer, Statistical and Mathematical Sciences, North-West University, Potchefstroom Campus, Potchefstroom, South Africa

[†]Corresponding author: (**Fellow of the Operations Research Society of South Africa**), School of Computer, Statistical and Mathematical Sciences, North-West University, Potchefstroom Campus, Potchefstroom, South Africa, email: Hennie.Kruger@nwu.ac.za

The 2014 Global Information Technology Report [36] issued by the World Economic Forum rated the quality of the educational system in South Africa as 146th out of 148 countries; the quality of mathematics and science education in South Africa was rated as last out of the 148 countries. The 2015 report [37] evaluated only 143 countries and in this report, the quality of the educational system in South Africa was rated as 139th out of the 143 countries. The quality of mathematics and science education was still rated as last out of the 143 countries. These poor statistics seem to be the norm and are confirmed by other similar studies, for example, in a Swiss study on the competitiveness of 61 countries, the “science in schools” in South Africa is ranked 59th and the overall South African educational system 60th [18]. Another alarming statistic is that of the 1.2 million children enrolled in Grade 1 in 2001, less than half (44%) reached Grade 12 [16].

There is also a general agreement between South African media news reports and academic studies that problems exist (and persist) in the South African school system. Media reports frequently highlight problems such as poor school performance [17, 19], more general support needed to assist underperforming schools [10], not enough financial resources allocated to schools [5], and the low classification of schools according to their performance [23]. Formal research projects that focus on the efficiency and effectiveness of schools are conducted continuously. Examples include those of Assan & Lumadi [2] (effectiveness of induction programmes for school principals), Spaul [29] (inequalities in primary schools), Spaul [30] (quality of education) and Taylor [33] (indicators of effective school management). Although a few years have already passed since completion of the work of Mouton *et al.* [24], it still provides a relevant and excellent exposition of challenges that are facing the South African education system. It also provides recommendations on the abatement of the identified challenges.

From these brief introductory comments, it is clear that the South African educational system finds itself in a rather dismal position. Taylor [33] elegantly summarises the problem of school efficiency as follows:

School functionality or efficiency [in South Africa] remains something of a “black box”: resources flow into the box and differential outcomes emerge, yet little is known or can be proven about what occurs within the box to determine the outcome.

Despite the general concern over the South African school system and the many research studies that have been conducted in this area, there are surprisingly few or no studies that utilise mathematical models or techniques to analyse the performance of schools in South Africa. The well-known data envelopment analysis (DEA) models are examples of mathematical modelling techniques that are specifically suited to measure efficiency of different decision-making units. These types of models have not yet been used formally in a South African schools context, although they are regularly applied in school performance studies overseas [1, 4, 11]. The models are not restricted to schools only but are also employed frequently to evaluate other educational institutions (*e.g.* universities) [15, 25, 35].

Against this background and considering the apparent lack of the use of mathematical models in the South African educational system, this paper proposes a mathematical approach to assess the performance and efficiency of schools in South Africa. The proposed framework employs a class-ranking technique that implements Pareto optimality principles

and makes use of a DEA-based model that disregards input variables and utilises outputs only. To demonstrate the implications, feasibility and usefulness of the proposed approach, the mathematical models have been applied to 54 schools in one of the four regional municipal districts in the North West province of South Africa. The results that have been obtained indicate that the suggested methodology and models do indeed contribute to the monitoring of schools and will certainly assist with achieving specific educational goals pertaining to effectiveness and efficiency. Furthermore, because the proposed models provide for the categorisation of schools in different classes and the simultaneous development of stepwise improvement plans, they are well aligned with some of the overall plans of the South African educational departments, for example the implementation of categories of schools to assist with the identification of weaknesses [23] and the use of (numerical) data to unlock the potential of the school system [12].

The remainder of the paper is organised as follows. Section 2 presents a brief background to schools in the North West province where the research was conducted. The class-ranking conceptual framework is described in §3, whereas §4 presents the application of the suggested techniques and models. Finally, §5 presents a discussion of the results and the paper is concluded in §6 with some final remarks.

2 Background

The South African National Department of Education consists of two ministries: Basic Education (childhood development centres and primary and secondary schools) and Higher Education and Training (tertiary institutions). The two ministries are each headed by a minister on national level and each of the nine provinces in South Africa has its own Provincial Department of Education. The central government provides a national framework for school policies, whereas the provincial governments are responsible for administering and implementation of the policies. At school level, public schools are governed by school governing bodies that normally consist of different representative members such as the school headmaster, parents, educators, other staff, and co-opted members. In secondary schools, learners may also form part of the governing body. The educational system is regulated by the *South African Schools Act 84* of 1996 and education is compulsory for all children, from the age of 7 (grade 1) to the age of 15 (grade 9). Although it is not compulsory, the majority of learners also attend grades 10 to 12 [14].

The 2015 school statistics for the nine provinces in South Africa indicated that there were 12 814 473 learners who attended 25 691 ordinary schools (public and independent schools). A total of 416 093 teachers were employed in the nine provinces. Table 1 presents a breakdown of the overall school statistics per province.

The proposed models in this study are applied to schools in the North West province, which consists of four main regional municipal districts called Bojanala Platinum (East), Dr Ruth Segomotsi Mopati/Bophirima (West), Ngaka Modiri Molema (Central) and Dr Kenneth Kaunda (South). The map in Figure 1 shows the geographical location of the four municipal regions in the North West province.

Due to data requirements and other cost considerations (see also §4.1 for additional expla-

Province	Number of learners	Number of educators	Number of schools	Number of grade 12 learners
Eastern Cape	1 953 397	64 256	5 727	89 322
Free State	682 704	23 661	1 332	35 545
Gauteng	2 262 319	79 354	2 780	120 838
KwaZulu-Natal	2 881 518	86 493	6 137	175 175
Limpopo	1 753 734	55 930	4 045	105 251
Mpumalanga	1 079 280	35 153	1 862	56 149
Northern Cape	290 139	9 162	581	12 498
North West	813 873	25 126	1 544	34 392
Western Cape	1 079 509	36 958	1 683	58 060
South Africa	12 814 473	416 093	25 691	687 230

Table 1: School statistics of all grades per province (2015) [15].

nations), this study only considers secondary schools (grades 8 to 12) that offer physical science and mathematics at grade 12 level in the Dr Kenneth Kaunda regional municipality. This selection resulted in a usable dataset (based on 2013 school statistics) of 54 schools [13]. Descriptive statistics for the selected dataset are shown in Table 2. Section 4.1 presents more detail on the selected dataset used in this study. The following section will discuss the class ranking framework to which the dataset was applied.

Description	Total
Number of schools	54
Total number of learners	45 287
Number of grade 12 learners	5 241
Number of grade 9 learners	12 006
Total number of learners taking mathematics	1 999
Total number of learners taking physical science	1 783
Total number of learners who passed grade 12	4 337

Table 2: Descriptive statistics (2013) of the dataset used in this study [13].

3 Class-ranking conceptual framework

The class-ranking concept that is proposed in this paper stems from the Pareto optimality concept where decision-making units, indicated as DMUs (schools in this paper), are stratified into classes of different levels. The technique was suggested by Barr *et al.* [3] and is often referred to as “peeling the DEA onion”. The basic premise behind this approach is the separation of DMUs into a series of nested efficient-frontier layers that provide a different efficiency metric as well as an explanation for those DMUs that are categorised as inefficient. The method goes hand in hand with the principle of context-dependent data envelopment analysis [27] that proves to be useful in situations where every DMU needs to focus on its own, inherently strong areas to develop a distinct character. In these cases, the conventional way of aggregating scores of variables to form a composite ranking index is inappropriate and decision makers may rather prefer the concept of dominance, for example a DMU that is Pareto optimal if it performs better in at least one aspect and

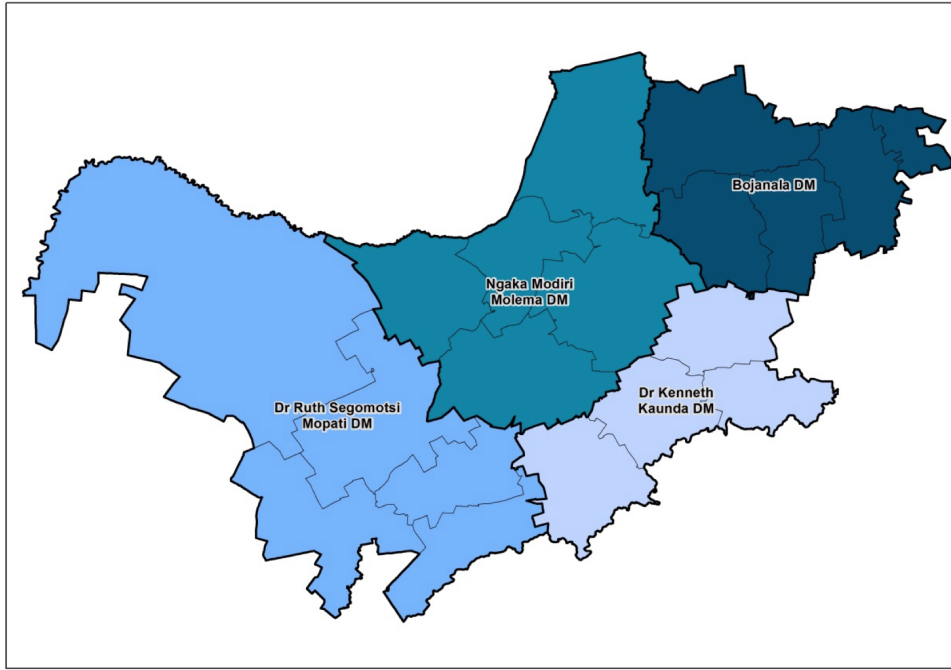


Figure 1: The four municipal districts in the North West province.

at least the same in all other aspects than any other DMU. The concept of dominance and Pareto efficiency, in the context of traditional data envelopment analysis models, is explained by Thanassoulis [34] as: Consider a homogenous set of decision-making units that consume a number of inputs to secure a number of outputs. For an *output-oriented* DEA model, a DMU is Pareto efficient if it is not possible to raise any one of its output levels without lowering at least another one of its output levels and/or without increasing at least one of its input levels. More formally, it can be defined as: Let y_{rj} ($r = 1, \dots, s$) be the output levels secured by DMU_j and x_{ij} ($i = 1, \dots, m$) the levels of input consumed. DMU_o is Pareto efficient if there exists no observed or feasible in principle DMU_j ($j \neq j_o$) such that $y_{r'j} > y_{r'j_o}$ for some r' and $y_{rj} > y_{rj_o}$ for all $r \neq r'$ while $x_{ij} < x_{ij_o}$ for all i . Pareto efficiency for input-oriented models is defined in a similar fashion by using inputs instead of outputs.

The concept of dominance and Pareto optimality can be illustrated by the following simple example (adapted from [20]). Consider six DMUs (labelled A to F) that are being evaluated by using two output variables (criteria). The values of the two outputs for each DMU are summarised in Table 3.

DMU	A	B	C	D	E	F
Output 1	1	4	5	3	4	3.1
Output 2	5	4	1	4	2	3.1

Table 3: The values for the two outputs for each DMU.

Based on the output values and assuming that the importance of the two output measures

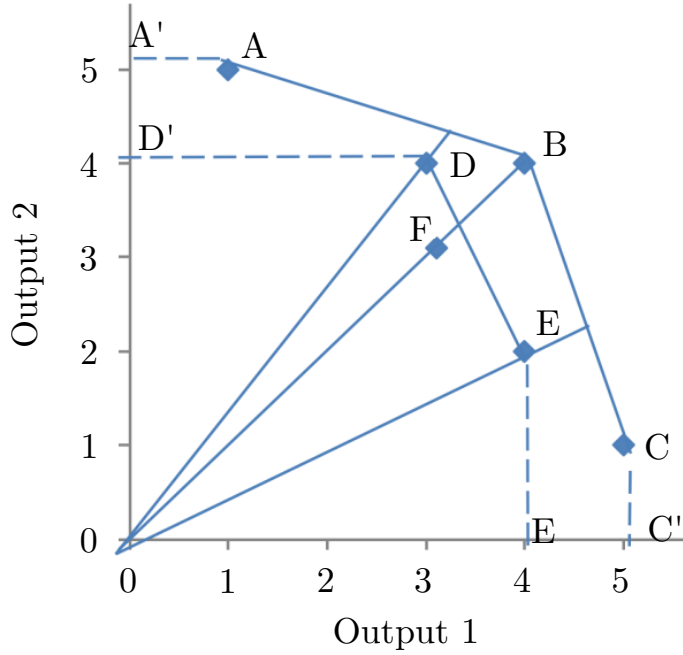


Figure 2: Dominance relationship of the six DMUs.

is irrelevant, DMU_B performs best with a total score of 8. Second best is DMU_D (with a total score of 7), followed by DMU_F (with a total score of 6.2). The remaining three DMUs (A, C and E) are tied in fourth place with a total score of 6 each. It should be noted that assigning different weights to the two criteria would result in a different ranking. The six DMUs and their associated output values are depicted in Figure 2.

To illustrate the dominance principle, consider the case where the two criteria are incomparable – this is equivalent to seeking for dominating DMUs. The dominating DMUs will belong to a class of the same level, whereas the dominated DMUs will belong to a class of a lower level. With reference to Figure 2, the DMUs A, B and C are the dominating DMUs and are also Pareto optimal. These three DMUs are therefore in the same class and on a higher level than the other DMUs. The other DMUs (D, E and F) are dominated by DMU_B and are therefore categorised in lower classes at lower levels. According to Kao & Lin [20], the dominance relationship can be identified graphically by connecting the DMUs on the northeast frontier with all the other DMUs inside the frontier. This can be seen from Figure 2 where the frontier is composed of the linear line segment $A'ABCC'$, with the Pareto optimal DMUs those on the frontier (A, B and C) and the dominated DMUs (D, E and F) those inside the frontier. Repeating this process will then result in a set of DMUs that can be divided into different levels of efficient frontiers. There are three different frontiers in Figure 2, namely $A'ABCC'$, $D'DEE'$ and F. Removing the original efficient frontier ($A'ABCC'$), the remaining inefficient (dominated) DMUs will form a new, second-level efficient frontier ($D'DEE'$). If this new, second-level frontier is removed, a third-level efficient frontier is formed (in Figure 2 this is the frontier consisting only of DMU_F) and so on until no DMUs are left. This process demonstrates the “onion peeling” concept with the different frontiers providing a class ranking as well as a means for context evaluation

of the various DMUs. The class ranking in this example, based on Pareto optimality, is therefore given as Class 1 (DMUs A, B and C); Class 2 (DMUs D and E); and Class 3 (DMU_F).

The process of generating non-dominated sets of DMUs (*i.e.* the nested frontiers) graphically would become difficult or impossible with more than two outputs and a large number of DMUs. A more general and efficient method to generate the required non-dominated sets of DMUs is to construct a context-dependent DEA model as proposed by Seiford & Zhu [27]. This model is formulated as follows:

Suppose $DMU_j (j = 1, \dots, n)$ produces a vector of outputs $\mathbf{y}_j = (y_{1j}, \dots, y_{sj})$ by using a vector of inputs $\mathbf{x}_j = (x_{1j}, \dots, x_{mj})$. Let $\mathcal{J}^l = \{DMU_j; j = 1, \dots, n\}$ be the set of all n DMUs and l indicating the current level of the nested efficiency frontiers. Define $\mathcal{J}^{l+1} = \mathcal{J}^l - \mathcal{E}^l$ where $\mathcal{E}^l = \{DMU_k \in \mathcal{J}^l | \phi^*(l, k) = 1\}$ and $\phi^*(l, k)$ is the optimal value to the following linear programming model, namely

$$\phi^*(l, k) = \max_{\lambda, \phi(l, k)} \phi(l, k) \quad (1)$$

subject to

$$\sum_{j \in F(\mathcal{J}^l)}^n \lambda_j \mathbf{y}_j \geq \phi(l, k) \mathbf{y}_{k'}, \quad (2)$$

$$\sum_{j \in F(\mathcal{J}^l)}^n \lambda_j \mathbf{x}_j \leq \mathbf{x}_k, \quad (3)$$

$$\lambda_j \geq 0, \quad (4)$$

$$j \in F(\mathcal{J}^l), \quad (5)$$

where $(\mathbf{x}_k, \mathbf{y}_k)$ represents the input and output vectors of DMU_k and $j \in F(\mathcal{J}^l)$ means that $DMU_j \in \mathcal{J}^l$, *i.e.*, $F(\cdot)$ represents the correspondence from a DMU set to the corresponding subscript index set. Non-negative weights λ_i are associated with individual inputs i or outputs i .

In the case of $l = 1$, the formulated model (constraints (1) to (5)) equates to the original output-oriented CCR model [9], and DMUs in the set $\mathcal{E}^1$ define the first-level efficient frontier — the set of Pareto optimal DMUs (non-dominated) will be ranked into the first class. When $l = 2$, the model solution will give the second-level efficient frontier after the exclusion of the first-level Pareto optimal DMUs. This process is then repeated until all different levels of efficient frontiers have been identified. For every l , $\mathcal{E}^l$ will contain the Pareto optimal set of DMUs (class ranking) for the l^{th} level efficient frontier. This process of class ranking a group of DMUs, based on the nested efficient frontier procedure, can formally be expressed in the following steps [27]:

Step 1: Set $l = 1$. Evaluate the entire set of DMUs $\mathcal{J}^l$ by the model (1) to (5) to obtain the first-level efficient DMUs set $\mathcal{E}^l$ (the first-level efficient frontier).

Step 2: Exclude the efficient DMUs from future DEA runs using $\mathcal{J}^{l+1} = \mathcal{J}^l - \mathcal{E}^l$ (if $\mathcal{J}^{l+1} = \phi$, then stop).

Step 3: Evaluate the new subset of “inefficient” DMUs $\mathcal{J}^{l+1}$ by model (1) to (5) to obtain a new set of efficient DMUs $\mathcal{E}^{l+1}$ (the new efficient frontier).

Step 4: Let $l = l + 1$. Go to *step 2*.

With reference to the illustrative example that is graphically depicted in Figure 2, the application of this algorithm would yield a three-class stratification with A, B and C in the first, D and E in the second and F in the third class. DMUs in the same class are of the same rank. Comparing these rankings to the initial equal-weight rankings, two differences can be noted: First, A and C, which have the lowest total score (6 each), are now classified into the first class. Second, F, which has the third-highest score, has now been classified in the lowest class. This is due to the equal-weighted method that favours DMUs that perform equally well in both criteria, whereas the dominance method favours DMUs that perform extraordinarily well in specific criteria [20].

In this paper, a slightly different model was implemented to generate non-dominated sets of DMUs (schools). The model formulation is based on the basic DEA principles but utilises only output variables — this approach was suggested by Kao & Lin [20]. Considering only output variables seems to be more appropriate in the context of schools, as there are normally significant differences in resources (inputs) available for different schools. The motivation for using outputs only will be further expanded on in §4 and §5.

The model (without inputs) is formulated with the objective to

$$\text{maximise } E_k = \sum_{j=1}^m u_j y_{kj} \quad (6)$$

subject to

$$\sum_{j=1}^m u_j y_{ij} \leq 1 \quad \text{for } i = 1, \dots, n, \quad (7)$$

$$u_j \geq \varepsilon \quad \text{for } j = 1, \dots, m, \quad (8)$$

$$\varepsilon > 0, \quad (9)$$

where y_{ij} represents output j (m outputs) for DMU_i (n DMUs), u_j indicates the weight of output j and ε is a small positive number used to restrict a DMU from ignoring unfavourable criteria.

Solving this model would result in DMU_k selecting the most advantageous weights to calculate a composite index E_k . If $E_k = 1$, then DMU_k is Pareto optimal, otherwise it is Pareto non-optimal. The four-step algorithm presented above and the implementation of the output-only model (constraints (6) to (9)) are now employed to generate the Pareto optimal sets of DMUs. This model and the algorithm were also used successfully in other similar studies [3, 20, 27, 31].

Solving the dual formulation of the primal model in equations (6) to (9) will provide more information regarding target values for a specific DMU and is therefore more appropriate to use.

The dual model [20] is formulated as

$$\text{minimise } E_k = \sum_{i=1}^n \lambda_i - \varepsilon \sum_{j=1}^m s_j \quad (10)$$

subject to

$$\sum_{i=1}^n y_{ij} \lambda_i - s_j = y_{kj} \quad \text{for } j = 1, \dots, m, \quad (11)$$

$$\lambda_i \geq 0 \quad \text{for } i = 1, \dots, n, \quad (12)$$

$$s_j > 0 \quad \text{for } j = 1, \dots, m. \quad (13)$$

Although the basic solutions of the primal and dual models are the same, the targets for a specific DMU can be calculated more readily by using the output from the dual formulation. To achieve this, let $\theta = \sum_{i=1}^n \lambda_i$. The targets for DMU_k is then given by

$$(y_{kj} + s_j)/\theta \quad \text{for } j = 1, \dots, m. \quad (14)$$

The multi-stage class-ranking model will therefore provide intermediate goals for each DMU (from one class to the next); this is in contrast to the conventional one-stage DEA model that only provides an ultimate goal of getting directly to the initial efficient frontier (the top class). For a DMU belonging to class f , the $f - 1$ targets in the previous $f - 1$ stages can be calculated. Each of these targets will then show the improvement that is needed by a DMU in each category to become a member of the corresponding class.

In the context of this study, the application of the algorithm and the associated models would provide a ranking of schools under incomplete criteria (output variables). Furthermore, the dual formulation presented above will also enable the construction of intermediate goals for a school to become efficient in stages (*i.e.* moving from a lower frontier to the next higher frontier). These will be further explained in §4, where the models will be applied to the schools data set.

4 Model application

This section presents the application of the mathematical class-ranking technique to schools in the North West province of South Africa. The aim is to provide an efficiency assessment of the schools and, following a synopsis of the school dataset, two applications will be offered. First, an application of the well-known CCR DEA model (utilising both input and output variables) will be given for comparative reasons, followed by a description of the “output-only” school-ranking model as detailed in §3.

4.1 The school dataset

The selection of input and output variables in the school data set was completely dependent on the availability of data. It should therefore be noted that different (or additional)

input and output variables may have been reported in similar studies in the literature. For example, a study by Lovell *et al.* [21] that was published in Charnes *et al.* [8] made use of variables such as physical facilities and extra-curricular activities. This type of data was not available in the North West province school system and could only be obtained at a very (and unacceptable) high cost by visiting each school physically to record the data through a process of inspection and/or interviews. Furthermore, to comply with the properties of valid data, the data used in the analysis were limited to one of the four main municipality districts (Dr Kenneth Kaunda district) in the North West province (see §2 for a brief description of the different municipalities). Reasons for this limitation include the following:

- *Access:* There was easy access to the Department of Education for this specific municipality district.
- *Availability:* The data for this municipality district were readily available.
- *Completeness:* A lack of or incomplete data for the other districts mean that they would be unsuitable for the study unless a huge amount of cost and effort would be invested to obtain proper data.
- *Authorisation:* Permission to include schools from other districts could not be obtained.
- *Sufficiency:* The number of schools in the selected municipality district was sufficient to apply the proposed mathematical models and perform an efficiency assessment.

To ensure a homogeneous dataset, it was further decided to exclude primary schools and use only secondary schools (grades 8 to 12) in the study. Additionally, only those secondary schools offering mathematics and physical science were considered. This decision was based on the importance of mathematics and physical science as subjects. Following an initial data analysis (to remove records with missing data values and to perform appropriate data transformations), a useable dataset of 54 secondary schools was obtained. Descriptive statistics for the 54 schools were presented in Table 2 (§2). From the identified dataset, the following input and output variables were selected.

Input variables:

- *Number of educators:* This refers to the number of teachers in a school.
- *Number of subjects:* This is the number of different subjects offered by a school.
- *Learner/educator ratio:* This ratio indicates the number of learners per teacher in a school.
- *Qualification:* The qualification variable refers to the number of teachers with an honours or higher degree at a school.
- *Total subsidy per learner:* This represents school fees and a departmental subsidy per learner.

Output variables:

- *Grade 12 pass rate:* The grade 12 pass rate is expressed as a percentage for each school.
- *Grade 12 mathematics average mark:* Expressed as a percentage, this represents the average mark obtained in mathematics (grade 12) for each school.
- *Grade 12 physical science average mark:* Expressed as a percentage, this represents the average mark obtained in physical science (grade 12) for each school.
- *School pass rate:* This variable represents the overall pass rate in all grades for a school and is expressed as a percentage.
- *Grade 9 pass rate:* This represents the pass rate (percentage) for grade 9 learners for each school. The grade 9 pass rate was deemed to be important, as grade 9 is a national examination (the same as for grade 12).

The complete dataset for the 54 schools and the data values are detailed in Appendix A.

4.2 The CCR DEA model

The aim of this study is to demonstrate the applicability of a class-ranking model (based on output variables only) and an associated context-dependent model to assess the efficiency of secondary schools. For comparative purposes and to highlight specific advantages of the class-ranking approach, a brief presentation of the basic CCR DEA model results, using the input and output variables that have been identified in §4.1, will also be presented.

The formulation of the basic CCR model applied to the school data set is as follows [9]. Assume that a set of DMUs needs to be evaluated. Each DMU_j ($j = 1, \dots, n$) uses m inputs x_{ij} ($i = 1, \dots, m$) and generates s outputs y_{rj} ($r = 1, \dots, s$). To measure the technical efficiency of DMU_o , the following fractional programming model may be solved. The objective is to

$$\text{maximise } e_o = \frac{\sum_{r=1}^s u_r y_{ro}}{\sum_{i=1}^m v_i x_{io}} \quad (15)$$

subject to

$$\sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \leq 0 \quad \text{for } j = 1, \dots, n, \quad (16)$$

$$u_r, v_i \geq \varepsilon \quad \text{for } i = 1, \dots, m \text{ and } r = 1, \dots, s, \quad (17)$$

where ε is a non-Archimedean value designed to enforce strict positivity on the importance variables u_r and v_i . This model formulation provides for constant returns to scale.

The theory of fractional programming suggests that the fractional programming model above can be converted to a linear programming model by implementing appropriate variable substitutions [7]. Details of the conversion and variable substitutions fall outside the scope of this paper and only the final linear programming model that was utilised in this study is given. The objective is to

$$\text{maximise } e_o = \sum_{r=1}^s \mu_r y_{ro} \quad (18)$$

subject to

$$\sum_{i=1}^m v_i x_{i0} = 1 \quad (19)$$

$$\sum_{r=1}^s \mu_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \leq 0 \quad \text{for } j = 1, \dots, n, \quad (20)$$

$$\mu_r, v_i \geq \varepsilon \quad \text{for } i = 1, \dots, m \text{ and } r = 1, \dots, s. \quad (21)$$

Solving the linear programming model (in equations (18) to (21)) by using the input and output values of the 54 schools (see Appendix A), the following results (Table 4) were obtained:

School	Efficiency rating	Reference set	School	Efficiency rating	Reference set
School 1	1.000		School 28	0.918	5, 16, 22, 29
School 2	1.000		School 29	1.000	
School 3	0.956	1, 5, 9, 24, 25	School 30	0.578	5, 18, 20, 29
School 4	0.926	2, 8, 9, 10, 20	School 31	1.000	
School 5	1.000		School 32	0.921	5, 22, 29
School 6	0.830	1, 2, 5, 9, 24, 25	School 33	0.693	5, 12, 18, 20, 24, 29
School 7	0.928	2, 5, 8, 9	School 34	0.908	19, 20, 29
School 8	1.000		School 35	1.000	
School 9	1.000		School 36	0.803	2, 5, 29
School 10	1.000		School 37	0.924	20, 22, 29
School 11	1.000		School 38	0.896	5, 24, 35
School 12	1.000		School 39	0.789	5, 9, 20, 24
School 13	0.865	1, 2, 29	School 40	0.812	5, 16, 22
School 14	0.880	1, 2, 10, 20	School 41	0.997	11, 19, 29, 46
School 15	1.000		School 42	0.760	2, 5, 20
School 16	1.000		School 43	0.867	2, 20, 25, 29
School 17	0.975	5, 15, 16, 24, 25, 29	School 44	0.842	5, 16, 22, 29
School 18	1.000		School 45	0.662	5, 16, 18, 24, 29
School 19	1.000		School 46	1.000	
School 20	1.000		School 47	0.780	5, 9, 24, 25
School 21	0.995	5, 16, 24, 35	School 48	0.579	2, 5, 29
School 22	1.000		School 49	0.998	2, 5, 22, 29
School 23	0.897	2, 5, 29	School 50	0.804	5, 16, 24, 35
School 24	1.000		School 51	0.646	2, 5, 20, 22, 29
School 25	1.000		School 52	0.894	5, 22
School 26	0.953	12, 18, 20, 24, 29	School 53	0.948	5, 20, 22, 35
School 27	0.863	5, 9, 12, 20, 24	School 54	0.509	5, 29, 35

Table 4: Results of the CCR model.

From Table 4 it can be seen that, given the set of input and output values, 37% (20 schools) of the schools were evaluated as efficient (with an efficiency rating of 1). The remainder of the schools are relatively inefficient. By using the identified reference sets (column 3 in Table 4), specific targets for the inefficient schools can be calculated to improve their performance to a 100% level of efficiency. A reference set for an inefficient school can be defined as the set of efficient schools with which the inefficient school is most comparable. A more comprehensive explanation of reference sets, how they are determined and how

they can be employed to calculate targets do not form part of the scope of this paper — details on this can be found in Metzger [22] and Ragsdale [26]. Target setting for inefficient schools was not performed, as the goal of this study is to focus on class ranking of schools and the construction of a stepwise improvement plan (progress from one class to the next higher class) rather than making recommendations on how to become 100% efficient immediately. Such a recommendation would be inappropriate (and useless) in the context of schools, as many of the inefficient schools simply do not have the required resources to become as efficient as the schools in their reference sets.

The results of the CCR model in Table 4 revealed no specific trend or clustering of both efficient and inefficient schools. The schools rated as “efficient” occur in urban areas, rural areas and townships, and no definite conclusion in terms of groups of schools are possible. Other than providing a list of efficient and inefficient schools, the results seem to have no additional significant value. One of the reasons for this may be the limiting set of inputs that were used. As explained earlier, there may have been other input variables that were inaccessible for this study that would have had an impact on the schools assessment. The result of the CCR model in Table 4 will be used to discuss and contrast the results of the proposed class-ranking model that utilises output variables only.

4.3 The class-ranking model applied to the schools dataset

The main focus of this study is to show the feasibility of a class-ranking model that can be used for school efficiency assessment and is based on output variables only. The decision to utilise only output variables is centred on the general assumption of unfair distribution of resources (inputs) to schools particularly in the South African context. To achieve the objective of class ranking of schools and at the same time produce a realistic improvement plan (from one class to the next higher class) for those schools ranked in lower classes, the model formulated in (10) to (13) is solved repeatedly according to the four-step algorithm that has been detailed in §3. Intermediate targets for inefficient schools are calculated simultaneously by using formula (14), which has also been explained in §3. The five output variables identified in §4 were used. Applying the algorithm and the proposed model, two types of outputs were produced: a class ranking of the 54 schools; and intermediate targets for those schools that are not ranked in the top class. The following two subsections will describe the two types of outputs obtained, followed by a discussion and some final comments in §5.

4.3.1 The class-ranking results

The application of the four-step algorithm and utilisation of the output-only model resulted in the classification of schools as detailed in Table 5. From this table, the following is observed: A total of 11 different classes, corresponding to 11 nested efficiency frontiers, were identified for the 54 schools under study. Class 1 (the top class) consists of 2 schools, class 2 has 4 schools, class 3 has 5 schools and so forth. The lowest class is class 11 and consists of only one school. The results also clearly indicate the Pareto optimality principle, as schools in higher classes dominate those categorised in lower classes.

It is interesting to note that an analysis of the different school locations revealed that

School	Grade 12 pass rate	Mean grade 12 Maths	Mean grade 12 Science	School pass rate	Grade 9 pass rate	Class
School 1	100.0	57.0	59.7	100.0	100.0	1
School 3	100.0	61.8	61.2	98.6	100.0	1
School 2	100.0	46.3	36.7	100.0	100.0	2
School 4	100.0	59.1	59.4	97.4	96.0	2
School 6	100.0	56.2	55.4	96.8	97.0	2
School 8	98.0	54.0	60.6	95.0	95.0	2
School 5	98.0	58.7	47.9	96.8	93.0	3
School 7	97.0	51.1	59.5	96.0	94.0	3
School 10	100.0	49.2	54.6	94.2	89.0	3
School 12	86.0	39.8	40.3	90.4	96.0	3
School 18	85.0	30.1	34.0	78.8	100.0	3
School 9	96.0	49.4	50.6	94.4	86.0	4
School 13	94.0	40.9	41.4	88.8	92.0	4
School 11	95.0	39.9	46.7	91.2	88.0	5
School 14	92.0	46.7	46.9	82.8	64.0	5
School 15	88.0	31.4	48.4	81.2	81.0	5
School 19	96.0	40.3	45.9	73.4	54.0	5
School 24	63.0	46.2	46.6	68.6	85.0	5
School 25	96.0	44.7	47.4	68.0	53.0	5
School 31	90.0	47.8	42.4	64.0	31.0	5
School 16	92.0	38.5	46.4	80.6	64.0	6
School 17	87.0	34.6	43.5	79.6	76.0	6
School 20	90.0	40.4	38.9	72.0	46.0	6
School 21	87.0	34.5	46.5	70.8	75.0	6
School 22	95.0	28.6	34.4	70.2	25.0	6
School 29	82.0	25.0	29.9	66.4	79.0	6
School 38	74.0	44.2	36.5	59.4	57.0	6
School 23	57.0	21.0	25.9	69.4	68.0	7
School 26	72.0	33.8	32.9	68.0	59.0	7
School 27	63.0	28.9	39.8	67.2	67.0	7
School 28	87.0	28.7	37.4	66.4	61.0	7
School 30	72.0	25.9	19.2	65.8	69.0	7
School 32	88.0	35.3	26.2	63.0	54.0	7
School 35	87.0	33.6	36.0	60.6	41.0	7
School 39	76.0	34.8	41.6	58.6	57.0	7
School 49	94.0	31.1	34.1	47.4	34.0	7
School 33	62.0	28.9	30.6	62.2	63.0	8
School 34	86.0	21.0	35.1	61.2	30.0	8
School 37	84.0	27.6	31.9	60.2	48.0	8
School 41	70.0	32.0	38.0	58.4	19.0	8
School 43	89.0	31.0	34.6	56.2	36.0	8
School 46	73.0	23.6	28.1	54.6	61.0	8
School 47	66.0	25.6	37.8	53.2	39.0	8
School 36	63.0	17.4	24.5	60.4	54.0	9
School 40	76.0	29.1	34.4	58.6	29.0	9
School 42	69.0	27.5	31.7	56.6	41.0	9
School 44	80.0	29.1	31.8	54.8	37.0	9
School 45	52.0	24.5	27.2	54.8	50.0	9
School 50	74.0	29.7	34.5	47.4	37.0	9
School 52	86.0	20.4	30.3	42.6	2.0	9
School 48	72.0	23.5	25.6	53.0	26.0	10
School 51	70.0	25.6	29.8	44.6	2.0	10
School 53	74.0	23.4	29.0	34.6	2.0	10
School 54	46.0	21.2	14.1	30.4	3.0	11

Table 5: *Class-ranking results for the 54 schools.*

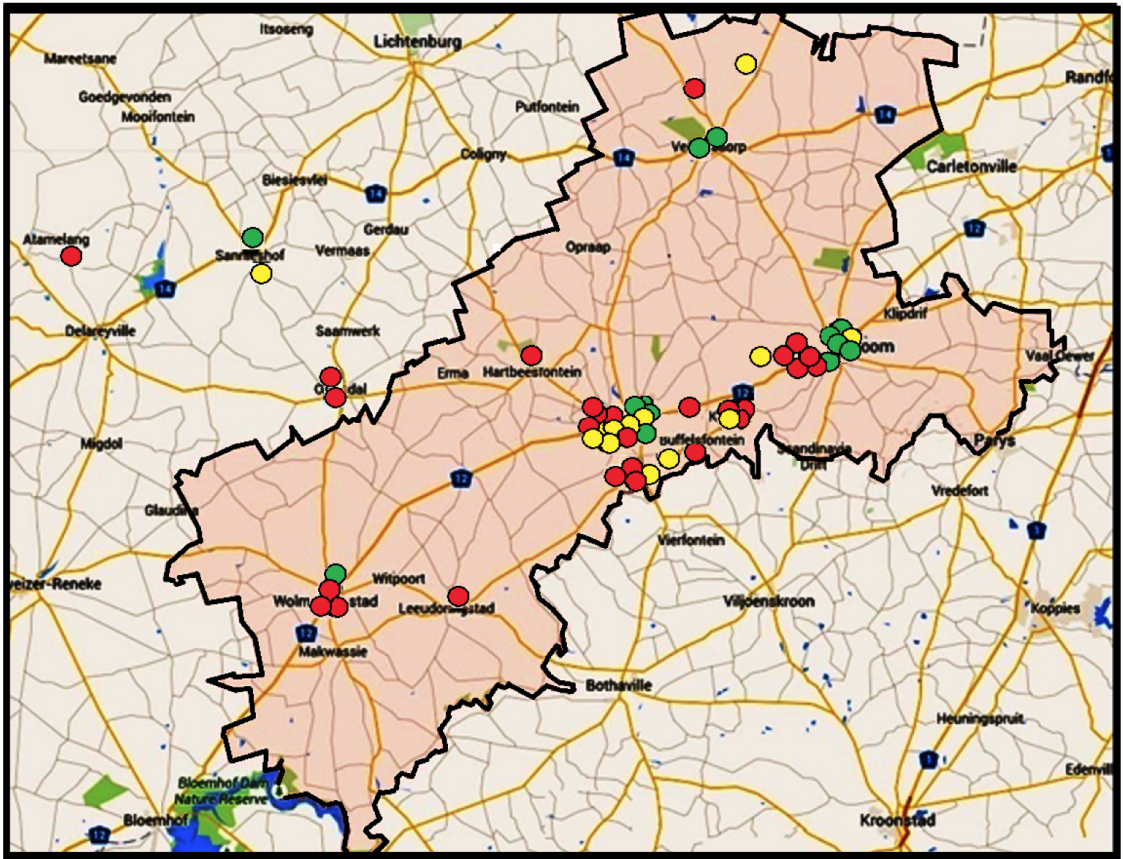


Figure 3: Ranking of the schools. [Figure can be viewed in colour in the electronic version, available at <http://orion.journals.ac.za/>.]

schools categorised in classes 1 to 4 are mainly urban area schools (traditionally advantaged schools); schools categorised in classes 7 to 11 are mostly rural, farm and township schools (traditionally disadvantaged schools); and schools categorised in classes 5 and 6 represent a mix of traditionally advantaged and disadvantaged schools. This observation confirms that the class-ranking model produced logically and intuitively correct and reliable results that are consistent with expectations.

The class-ranking results are graphically displayed on the map that is presented in Figure 3. The schools in the top 4 classes are labelled green, whereas schools in classes 5 and 6 are labelled yellow; red labels are used for the less efficient schools in classes 7 to 11. For clarity purposes, the two main urban areas in the region (Potchefstroom and Klerksdorp-Orkney-Stilfontein) are also shown separately in Figures 4 and 5. The maps visually confirms the distribution of schools in urban areas (green-labelled schools) and rural and township areas (yellow- and red-labelled schools).

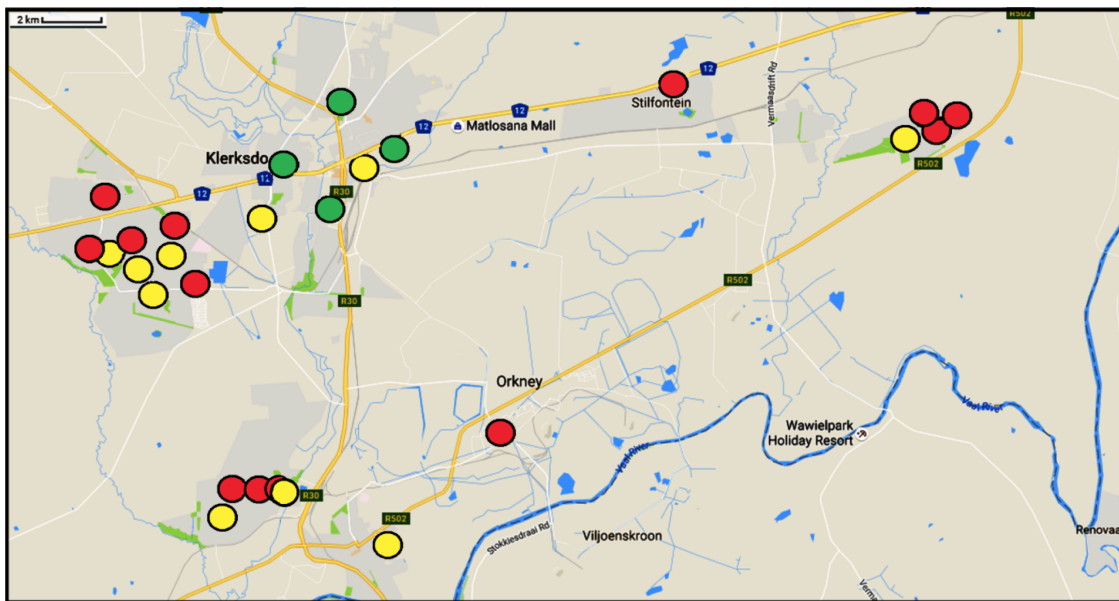


Figure 4: School rankings in the Klerksdorp-Orkney-Stilfontein area. [Figure can be viewed in colour in the electronic version, available at <http://orion.journals.ac.za>.]

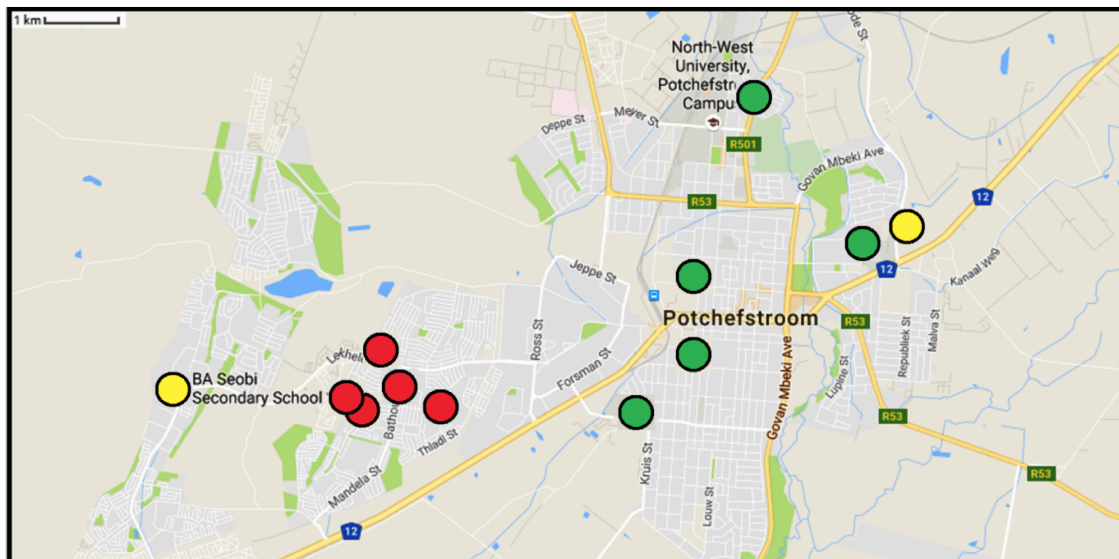


Figure 5: School rankings in the Potchefstroom area. [Figure can be viewed in colour in the electronic version, available at <http://orion.journals.ac.za>.]

4.3.2 Intermediate targets for inefficient schools

An added strength of the multi-stage class-ranking model is that by using formula (14), intermediate targets for inefficient schools can be calculated. These intermediate targets indicate requirements or improvements that are necessary to move from a current class to the next higher class. This is of particular interest, as inefficient schools do not normally have resources to progress immediately from a currently lower class to the top class (*i.e.* what a traditional CCR model would suggest); they rather try to improve their performance to the next higher class. The technique therefore provides a roadmap to schools to improve their efficiency in a more realistic and step-by-step manner. To illustrate the usefulness of this, consider the intermediate targets that were calculated for School 40 and are presented in Table 6.

Class level	Grade 12 pass rate	Mean grade 12 Maths	Mean grade 12 Science	School pass rate	Grade 9 pass rate
1	100.0	57.0	59.7	100.0	100.0
2	100.0	59.1	59.4	97.4	96.0
3	100.0	49.2	54.6	94.2	89.0
4	96.0	45.8	48.2	74.0	60.5
5	96.0	40.3	45.9	74.0	55.1
6	92.9	35.6	42.8	77.5	52.4
7	82.7	31.7	37.4	63.8	55.8
8	77.1	29.5	34.9	59.4	33.9
Current (9)	76.0	29.1	34.4	58.6	29.0

Table 6: *Intermediate targets for School 40.*

School 40, which is a school from a township area, was ranked in class 9 (Table 5). It would be unreasonable and impossible to expect from a township school in the North West province to improve from its current inefficient position in class 9 to the same level of the top performers in class 1 in a short period of time (as a typical CCR model would suggest). Resources for such drastic improvements in all output criteria are simply not available. A more feasible recommendation would be to try improving its performance only to the next higher class (class 8) and then progress step by step until a satisfactory level of efficiency has been reached.

To illustrate the above argument, consider the current grade 9 pass rate (29%) of School 40. Given the limited resources of schools, usually associated with an equally limited prospect of obtaining more resources, it is more realistic to try and improve the 29% pass rate to the next higher level of 33.9% in class 8 instead of aiming at the goal of a 100% pass rate (class 1). School 40 can more readily improve its performance from its current class 9 to the next higher class (class 8) by increasing the 5 output variables only marginally; for example, the grade 12 pass rate should only be increased from 76% to 77.1%. Once the output levels for a class 8 ranking have been achieved, the school can then proceed in trying to improve to the next higher level (class 7), as indicated in Table 6. The process is repeated over time until the school reaches a satisfactory level of performance. The intermediate targets thus provide a more realistic framework whereby schools can improve in a progressive manner to become more efficient over time.

Some points need to be noted regarding the intermediate target-setting approach. In some

instances, the level of a specific output variable may be lower in the next higher class — this is an indication that the school has already performed at a satisfactory level in that specific criteria. For example, for School 40, the school pass rate on level 6 (77.5%) is higher than the school pass rate on level 5 (74%). To move from class 6 to class 5, School 40 has already achieved a satisfactory level of performance in the school pass rate criteria and improvements are only necessary in the remaining 4 output criteria.

For smaller datasets and especially in cases where the range of data values are small, intermediate target values may also indicate which school, or schools, should be treated as a benchmark for the school under consideration. This is, however, not always true and in cases where a combination of schools form a benchmark, it may be difficult to determine the specific benchmark schools. In the context of schools (and specifically in this study), it is not important to know which schools are the benchmark schools. Once intermediate targets have been calculated, an inefficient school can proceed to work towards the targets. Due to the different resources available to schools, an inefficient school will not follow a specific benchmarking school (with different resources available) but would rather plan its performance and possible improvements according to its own unique situation and available resources. The purpose of benchmarking schools is therefore limited to the calculation of intermediate targets for output variables.

The intermediate targets are calculated at a specific point in time. Due to this “snapshot” approach, it is advisable to solve the models again when schools progress from one class to the next. Moving from one class to the next (especially if a number of schools have improved over time) imply significant changes to the data, which may result in new target values. Moreover, the output variables chosen for this study are not independent from each other; for example, a change in the grade 9 pass rate will automatically result in a change of the school pass rate. This may have an influence on other intermediate targets. Despite these potential changes, the intermediate targets at a specific point in time still provide a reliable picture of a school relevant to other schools. It also gives a good indication of what may be expected from a school. Furthermore, improvements in efficiency (in the context of schools) are something that can only be achieved in the longer term — the intermediate targets will therefore provide direction to inefficient schools in terms of shorter term goals and assist with decisions on the use of available resources.

5 Discussion and concluding remarks

To highlight and emphasise the specific advantages of the proposed class-ranking approach, this section briefly contrasts the results of the class-ranking technique in §4.3.1 and §4.3.2 with the results obtained from applying a traditional CCR DEA model in §4.2.

5.1 The CCR model

The traditional CCR DEA model is well suited to assess the efficiency of different decision-making units in a wide variety of application areas. This approach is also applicable to the evaluation of the efficiency of educational institutions and particularly to schools. A prerequisite, however, is that the decision-making units (schools) must operate in a ho-

mogenous environment. In the context of schooling systems in developing countries (and certainly in South Africa), the operating environment is generally not homogenous, especially in the area of resources (inputs) where significant differences may exist. The performance and results of a CCR model, which is based on both input and output variables, may therefore be debateable if used to evaluate school efficiency. The following arguments (based on the results in §4) may also be considered.

The data for input variables in a school context present certain challenges. The quality of school input data is often questionable, as data are provided by schools where there may be a lack of control over the accuracy and integrity of the data. This is in stark contrast with school output variable data (*e.g.* pass rates), which are normally subjected to specific checking and validation procedures and are published widely. Another challenge is the *availability* of input data. Data may simply not exist, not be available or may only be obtained through difficult, lengthy and costly procedures (*e.g.* personal visits to schools). For example, an important input to any school efficiency model would be the number of additional facilities available at the school (*e.g.* library, computer centre, music centre and sports facilities). This type of data was not available in the current study and had to be excluded from the model. The conclusion from this is that there may be a high probability that important input variables will be excluded in school models that utilise input variables.

The nature of an input/output model and the associated model properties such as in the CCR model is that all schools are compared to each other relatively — the concepts of different classes and nested levels of efficiency are ignored. This result in contradictory evaluations; for example, a school that is evaluated as 100% efficient by the CCR model (see School 11 in Table 4) is categorised as a class 5 school only (see School 11 in Table 5) when taking Pareto optimality principles into account. The opposite is also true: School 3, for example, is classified as a class 1 school (Table 5) but is rated by the CCR model as being only 95% efficient (Table 4). It seems as if it is more appropriate in the context of a schools assessment to analyse the different classes or groups of school's instead of comparing all schools directly with each other.

Closely related to the above properties of an input/output model is the ability of a CCR model to provide targets for inefficient schools through the identification of the so-called reference sets. These targets are, however, aimed at transforming schools from their current inefficient levels (which may be very low) to the same level of the toprated, efficient schools. This type of target setting is useful in many applications, but in the context of South African schools, the targets seem to be senseless and of little use. As stated earlier, inefficient schools simply do not have resources (inputs) to become 100% efficient over a relatively short period of time. It makes more sense to provide an improvement plan with intermediate targets that will enable a more gradual progress in efficiency. This is something that the proposed output-only class-ranking model can provide and is further highlighted in the following section.

5.2 The output-only class-ranking model

The proposed output-only class-ranking model that implements Pareto optimal principles provides an opportunity to analyse school performances in different classes. In addition,

an improvement plan to guide the performance of inefficient schools in a stepwise manner can be constructed. The results that have been obtained seem to be much more reliable and appropriate in the context of a school's assessment.

Some of the arguments in favour of an output-only class-ranking model include the unfair or unequal distribution of resources to different schools and the availability of output data. Schools in South Africa are mainly judged by their outputs, namely their pass rates [28]. It therefore makes sense to focus on output variables and to exclude the uneven distribution of inputs and their potential negative impact on modelling results. In addition to this, the availability of verified and published output data, such as pass rates, contributes to more reliable and better quality model solutions. The construction of an improvement plan that enables schools in lower classes to move to the next higher class is much more practical and makes more sense in the context of schools where the uneven distribution of resources prohibits large and drastic improvements. Furthermore, applying the class-ranking principle makes it easier to identify trends or clusters of schools performing at the same level. The results presented in §4.3.1 clearly indicates and confirms that the top performing schools are clustered into urban areas, whereas the lower classified schools are located mainly in rural and township areas; this type of information cannot be seen easily from the CCR model results in §4.2.

Based on the results of this study and the ensuing discussions, it seems permissible to recommend that the proposed output-only class-ranking model should seriously be considered when schools are evaluated for efficiency. Schools in South Africa are unique and the efforts to help and improve the schooling system should be dynamic and aimed at the different requirements of different schools. The poor schools in the North West province seem to be focused on survival, whereas the wealthier schools are focused on performance with far less concerns about resources. These differences make class ranking and a gradual improvement plan the ideal tools to assist in addressing efficiency problems in schools.

6 Conclusion

Existing studies rank the educational system in South Africa constantly as almost last when compared to other countries. Despite the existence and availability of mathematical models to perform school efficiency evaluations, the use of such models in the context of South African schools is almost non-existent. Efforts to measure the quality of school education in South Africa focus mainly on the comparison of different schools in different countries [6]; however, this type of study only leads to the conclusion that South African learners receive schooling of inferior quality in comparison with those of poorer countries [32].

To address the problem of school efficiency as well as the apparent lack of mathematical models used in South African school performance evaluations, a class-ranking technique for schools was proposed. Pareto optimality principles form the basis of the ranking technique and were implemented in conjunction with a mathematical model that utilises only school-output variables. The mathematical model is also capable of constructing an intermediate step-by-step improvement plan for inefficient schools. The proposed technique was implemented in the Dr Kenneth Kaunda municipal region in the North West province

of South Africa by using a dataset of 54 secondary schools. Results were contrasted with the output from an ordinary input/output DEA model and it was clear that the class-ranking model and the associated intermediate improvement plans are better suited in the context of South African schools. Main reasons to support this claim include the following:

- The use of output variables only delivers more reliable results because of the general unavailability and low reliability of input data.
- Due to uneven (unfair) distribution of resources to schools, it would be impossible for an inefficient school to improve immediately (over a short period of time) to the same level as the efficient schools (as recommended by the typical DEA models). The implementation of an intermediate and step-by-step improvement plan (as proposed in this paper) makes more sense and is more readily achievable, especially when the availability of resources are taken into consideration.
- The Pareto principle ensures that schools are categorised to enable class/category comparisons as opposed to comparing each school with all other schools.
- It is easier to identify trends (clusters) in school performances by using the class-ranking results.

References

- [1] AGASISTI T, 2013, *The efficiency of Italian secondary schools and the potential role of competition: a data envelopment analysis using OECD-PISA2006 data*, Education Economics, **21**(5), pp. 520–544.
- [2] ASSAN TEB & LUMADI MW, 2013, *Effectiveness of the induction programme for school principals in the North West Province*, Anthropologist, **15**(1), pp. 79–87.
- [3] BARR RS, DURCHHOLZ ML & SEIFORD L, 2000, *Peeling the DEA onion: Layering and rank-ordering DMUs using tiered DEA*, Southern Methodist University, Texas, USA, Technical Report.
- [4] BLACKBURN V, BRENNAN S & RUGGIERO J, 2014, *Measuring efficiency in Australian Schools: A preliminary analysis*, Socio-Economic Planning Sciences, **48**(1), pp. 4–9.
- [5] CAPAZORIO B, 2013, *Millions spent on science and maths education, but little change*, Saturday Star: 12, 22 June, Report.
- [6] CARNOY M, CHISHOLM L & CHILISA B, 2012, *The low achievement trap: Comparing schooling in Botswana and South Africa*, HSRC Press: Cape Town.
- [7] CHARNES A & COOPER WW, 1962, *Programming with linear fractional functionals*, Naval Research Logistics Quarterly, **9**(3), pp. 181–186.
- [8] CHARNES A, COOPER WW, LEWIN AY & SEIFORD LM, 1994, *Data envelopment analysis: Theory, methodology, and applications*, New York, USA, Springer Science & Business Media.
- [9] CHARNES A, COOPER WW & RHODES E, 1978, *Measuring the efficiency of decision making units*, European Journal of Operational Research, **2**(6), pp. 429–444.
- [10] DAVIS G, 2013, *Underperforming schools can change with support*, Cape Argus: 15, 3 September.
- [11] DE FIGUEIREDO JN, BARRIENTOS M & ANGEL M, 2012, *A decision support methodology for increasing school efficiency in Bolivia's low-income communities*, International Transactions in Operational Research, **19**(1), pp. 99–121.
- [12] DELL M & DELL S, 2013, *Success by numbers. How using data can unlock the potential of South Africa's R-12 public school system*, Michael and Susan Dell Foundation.

- [13] EMIS, 2013, *Education Statistics in South Africa: School Realities*, Education Statistics in South Africa. Pretoria: Department of Basic Education.
- [14] EMIS, 2015, *Education Statistics in South Africa: School Realities*, Education Statistics in South Africa. Pretoria: Department of Basic Education.
- [15] HIRAO Y, 2012, *Efficiency of the top 50 business schools in the United States*, Applied Economics Letters, **19(1)**, pp. 73–78.
- [16] HOLBORN L, 2013, *Education in South Africa: where did it go wrong?*, Retrieved June 24, 2014, from Good Governance Africa, Available from <http://gga.org/stories/editions/aif-15-off-the-mark/education-in-south-africa-where-did-it-go-wrong>.
- [17] JAMAL S, 2016, *Schools slide off the charts*, The Times: 4, 16 March.
- [18] JOFFE H, 2016, *SA nudges up competitiveness rankings*, Business Day: 1, 31 May.
- [19] JOHN V, 2012, *Poorly performing schools 'slip through cracks'*, Mail and Guardian: 18, 19 July.
- [20] KAO C & LIN PH, 2008, *Class Ranking of the Management Colleges in Taiwan*, **1(1)**, pp. 129–140.
- [21] LOVELL C, GROSSKOPF S, LEY E, PASTOR J, PRIOR D & EECKAUT P, 1994, *Linear programming approaches to the measurement and analysis of productive efficiency*, TOP: An Official Journal of the Spanish Society of Statistics and Operations Research, **2(2)**, pp. 175–248.
- [22] METZGER LM, 1994, *Operational auditing and DEA: measuring branch office efficiency*, Internal Auditing, **10**, p. 3.
- [23] MONAMA T, 2016, *Categorisation of schools identifies weak areas*, The Star: 6, 16 March.
- [24] MOUTON N, LOUW G & STRYDOM G, 2013, *Critical challenges of the South African school system*, The International Business & Economics Research Journal, **12(1)**, pp. 31–44.
- [25] NAZARKO J & ŠAPARAUSKAS J, 2014, *Application of DEA method in efficiency evaluation of public higher education institutions*, Technological and Economic Development of Economy, **20(1)**, pp. 25–44.
- [26] RAGSDALE CT, 2007, *Managerial decision modeling*, Thomson: South-Western, Mason, USA.
- [27] SEIFORD LM & ZHU J, 2003, *Context-dependent data envelopment analysis: measuring attractiveness and progress*, Omega, **31(5)**, pp. 397–408.
- [28] SOUTH AFRICA, 2013, *Education in South Africa*, Retrieved April 18, 2016, from SouthAfrica.info, Available from <http://www.southafrica.info/about/education/education.htm>.
- [29] SPAULL N, 2012, *Poverty & privilege: Primary school inequality in South Africa*, International Journal of Educational Development, **33(5)**, pp. 436–447.
- [30] SPAULL N, 2013, *South Africas Education Crisis: The quality of education in South Africa 1994-2011*, Centre for development & enterprise (CDE), Informing South African policy. Johannesburg: CDE., pp. 1–65.
- [31] TAKEDA E & SATOH J, 2000, *A data envelopment analysis approach to multicriteria decision problems with incomplete information*, Computers & Mathematics with Applications, **39(9)**, pp. 81–90.
- [32] TAYLOR N, 2008, *Whats wrong with South African schools?*, Proceedings of the 2008 Conference on What's Working in School Development Conference, Jet Education Services, Cape Town, p. 21.
- [33] TAYLOR S, 2011, *Uncovering indicators of effective school management in South Africa using the National School Effectiveness Study*, Stellenbosch Economic working papers: 10/11, Department of Economics and the Bureau for Economic Research at the University of Stellenbosch.

- [34] THANASSOULIS E, 2001, *Introduction to the Theory and Application of Data Envelopment Analysis: A Foundation Text with Integrated Software*, 2nd Edition, Massachusetts: Kluwer academic publishers.
- [35] TÓTH R, 2009, *Using DEA to evaluate efficiency of higher education*, Applied Studies in Agribusiness and Commerce, **3(3)**, pp. 79–82.
- [36] WEF, 2014, *The Global Information Technology Report 2014*, Rewards and Risks of Big Data, World Economic Forum.
- [37] WEF, 2015, *The Global Information Technology Report 2015*, ICTs for Inclusive Growth, World Economic Forum.

Appendix A: The complete dataset of 54 schools

Data for the output variables are indicated as percentages. Table 7 contains the data for Schools 1–24 while Table 8 contains the data for Schools 25–54.

	Outputs					Inputs				
	Grade 12 pass rate	Grade 12 maths average	Grade 12 science average	School pass rate	Grade 9 pass rate	Number of educators	Number of subjects	Learner/educator ratios	Qualifications	Total subsidy per learner
School 1	100.0	57.0	59.7	100.0	100.0	34.0	15.0	15.1	20.0	14475.0
School 2	100.0	46.3	36.7	100.0	100.0	9.0	14.0	9.8	5.9	7585.0
School 3	100.0	61.8	61.1	98.6	100.0	38.0	19.0	19.0	21.1	9736.2
School 4	100.0	59.1	59.4	97.4	96.0	49.0	21.0	19.7	17.9	8569.9
School 5	98.0	58.7	47.9	96.8	93.0	24.0	19.0	18.3	13.7	75.0
School 6	100.0	56.2	55.4	96.8	97.0	25.0	21.0	19.9	30.8	8075.0
School 7	97.0	51.1	59.5	96.0	94.0	57.0	22.0	19.9	18.3	7025.0
School 8	98.0	54.0	60.6	95.0	95.0	46.0	20.0	19.1	15.1	8905.8
School 9	96.0	49.4	50.6	94.4	86.0	14.0	16.0	13.6	18.2	5575.0
School 10	100.0	49.2	54.6	94.2	89.0	38.0	16.0	21.0	13.2	7995.0
School 11	95.0	39.9	46.7	91.2	88.0	14.0	14.0	30.0	18.5	4975.0
School 12	86.0	39.8	40.3	90.4	96.0	46.0	13.0	20.7	6.3	4675.0
School 13	94.0	40.9	41.4	88.8	92.0	26.0	15.0	21.2	23.4	8575.0
School 14	92.0	46.7	46.9	82.8	64.0	49.0	17.0	21.1	10.6	7775.0
School 15	88.0	31.4	48.4	81.2	81.0	33.0	17.0	23.5	20.8	75.0
School 16	92.0	38.5	46.4	80.6	64.0	39.0	15.0	29.4	16.7	75.0
School 17	87.0	34.6	43.5	79.6	76.0	33.0	16.0	22.5	13.7	75.0
School 18	85.0	30.1	34.0	78.8	100.0	34.0	13.0	30.7	12.1	75.0
School 19	96.0	40.3	45.9	73.4	54.0	26.0	13.0	32.7	9.5	150.0
School 20	90.0	40.4	38.9	72.0	46.0	17.0	14.0	28.2	0.0	150.0
School 21	87.0	34.5	46.5	70.8	75.0	50.0	16.0	26.4	13.4	75.0
School 22	95.0	28.6	34.4	70.2	25.0	28.0	14.0	27.4	4.8	75.0
School 23	57.0	21.0	25.9	69.4	68.0	26.0	14.0	18.9	14.9	75.0
School 24	63.0	46.2	46.6	68.6	85.0	29.0	13.0	33.9	9.1	75.0

Table 7: Dataset for Schools 1–24.

	Outputs					Inputs				
	Grade 12 pass rate	Grade 12 maths average	Grade 12 science average	School pass rate	Grade 9 pass rate	Number of educators	Number of subjects	Learner/educator ratios	Qualifications	Total subsidy per learner
School 25	96.0	44.7	47.4	68.0	53.0	30.0	12.0	36.6	15.8	125.0
School 26	72.0	33.8	32.9	68.0	59.0	32.0	13.0	37.2	4.3	150.0
School 27	63.0	28.9	39.8	67.2	67.0	40.0	15.0	28.7	8.1	575.0
School 28	87.0	28.7	37.4	66.4	61.0	36.0	16.0	26.3	14.8	75.0
School 29	82.0	25.0	29.9	66.4	79.0	14.0	10.0	23.6	13.6	75.0
School 30	72.0	25.9	19.2	65.8	69.0	34.0	20.0	35.4	16.4	125.0
School 31	90.0	47.8	42.4	64.0	31.0	30.0	14.0	34.8	10.5	150.0
School 32	88.0	35.3	26.2	63.0	54.0	31.0	16.0	38.2	9.7	75.0
School 33	62.0	28.9	30.6	62.2	63.0	36.0	16.0	32.8	9.7	125.0
School 34	86.0	20.9	35.1	61.2	30.0	34.0	13.0	32.6	8.6	125.0
School 35	87.0	33.6	36.0	60.6	41.0	16.0	14.0	31.7	0.0	75.0
School 36	63.0	17.4	24.5	60.4	54.0	31.0	13.0	19.2	16.1	425.0
School 37	84.0	27.6	31.9	60.2	48.0	42.0	13.0	28.1	5.7	125.0
School 38	74.0	44.2	36.5	59.4	57.0	44.0	17.0	32.2	8.7	75.0
School 39	76.0	34.8	41.6	58.6	57.0	67.0	19.0	25.8	12.1	725.0
School 40	76.0	29.1	34.4	58.6	29.0	40.0	15.0	28.3	12.5	75.0
School 41	70.0	32.0	38.0	58.4	19.0	19.0	12.0	30.9	17.2	175.0
School 42	69.0	27.5	31.7	56.6	41.0	52.0	16.0	25.7	3.8	175.0
School 43	89.0	31.0	34.6	56.2	36.0	20.0	13.0	39.2	25.0	125.0
School 44	80.0	29.1	31.8	54.8	37.0	36.0	15.0	31.1	12.8	75.0
School 45	52.0	24.5	27.2	54.8	50.0	24.0	15.0	30.0	13.8	75.0
School 46	73.0	23.6	28.1	54.6	61.0	9.0	11.0	49.3	11.1	75.0
School 47	66.0	25.6	37.8	53.2	39.0	40.0	14.0	32.9	13.5	150.0
School 48	72.0	23.5	25.6	53.0	26.0	34.0	16.0	34.6	28.8	150.0
School 49	94.0	31.1	34.1	47.4	34.0	28.0	14.0	25.9	7.5	125.0
School 50	74.0	29.7	34.5	47.4	37.0	35.0	16.0	29.6	9.1	75.0
School 51	70.0	25.6	29.8	44.6	2.0	25.0	17.0	30.1	7.4	150.0
School 52	86.0	20.4	30.3	42.6	2.0	33.0	16.0	31.4	21.3	75.0
School 53	74.0	23.4	29.0	34.6	2.0	40.0	13.0	21.9	4.0	75.0
School 54	46.0	21.2	14.1	30.4	3.0	19.0	15.0	32.4	10.6	75.0

Table 8: *Dataset for Schools 25–54.*