

Reflecting on an impact evaluation of the Grade R programme: Method, results and policy responses

Authors:

Marie-Louise Samuels¹
 Stephen Taylor^{2,3}
 Debra Shepherd³
 Servaas van der Berg³
 Christel Jacob⁴
 Carol N. Deliwe⁵
 Thabo Mabogoane⁶

Affiliations:

¹Early Childhood Development, Department of Basic Education, South Africa

²Researcher and Advisor, Department of Basic Education, South Africa

³Department of Economics, Stellenbosch University, South Africa

⁴Director in the Evaluation and Research Unit, Department of Planning, Monitoring and Evaluation, South Africa

⁵Strategic Planning, Research and Coordination, Department of Basic Education, South Africa

⁶Outcomes Facilitator, Outcomes 1 and 5, Department of Planning, Monitoring and Evaluation, South Africa

Correspondence to:
 Stephen Taylor

Email:
 taylor.s@dbe.gov.za

Postal address:
 Private Bag X895, Pretoria
 0001, South Africa

Dates:
 Received: 10 Apr. 2015
 Accepted: 15 July 2015
 Published: 31 Aug. 2015

Read online:



Scan this QR code with your smart phone or mobile device to read online.

This paper describes the expansion since 2001 of a public pre-school programme in South Africa known as 'Grade R', summarises the findings from an impact evaluation of the introduction of Grade R, discusses the policy recommendations flowing from the evaluation and reflects on the process of implementing the recommendations. The Grade R programme has expanded dramatically, to the point where participation is nearly universal. Although a substantial literature points to large potential benefits from pre-school educational opportunities, the impact evaluation reported on in this article demonstrated that the Grade R programme, as implemented until 2011, had a limited impact on later educational outcomes. Improving the quality of Grade R, especially in schools serving low socio-economic status communities, thus emerges as a key policy imperative. Recommended responses include professionalising Grade R teachers, providing practical in-service support, increasing access to appropriate storybooks, empowering teachers to assess the development of their learners, and improving financial record-keeping of Grade R expenditure by provincial education departments. The impact evaluation was initiated by the Department of Planning, Monitoring and Evaluation (DPME) and the Department of Basic Education (DBE), and was conducted by independent researchers. The move towards increased evaluation of key government programmes is important for shifting the focus of programme managers and policymakers towards programme outcomes rather than only programme inputs. Yet the process is not without its challenges: following a clear process to ensure the implementation of the lessons learned from such an evaluation is not necessarily straightforward.

Introduction

This paper reports on a recent impact evaluation of the Grade R programme in South Africa. Grade R is a single-year pre-school programme intended for children in the year before entering Grade 1. It is implemented at primary schools or at community-based early childhood development (ECD) sites.¹ This programme has been systematically introduced and expanded by the South African government with the intention of preparing children from low socio-economic status communities for primary schooling.

Goldman *et al.* (2015) describes the development of a new National Evaluation Policy Framework and the adoption of a National Evaluation Plan (NEP). Under this plan, the Department of Planning, Monitoring and Evaluation (DPME) in the Office of the Presidency works together with other government departments to evaluate key programmes and policies.

In 2011, the Grade R programme was selected as one of the first evaluations of the NEP. A team of researchers from the University of Stellenbosch was contracted to conduct an impact evaluation. This impact evaluation has now been completed, has been presented to the Cabinet and is publicly available. It is therefore an opportune time to reflect on its findings, on how the evaluation has been received by various stakeholders, and on how it is influencing policy and programme design.

The evaluation terms of reference were approved by the steering committee on 05 September 2012. The service provider was contracted through the DPME procurement process on 12 December 2012. The final evaluation report was approved by the steering committee on 12 June 2013. The Department of Basic Education (DBE) provided a management response to the evaluation on

1. The impact evaluation reported on in this paper focuses only on the Grade R programme as implemented in ordinary primary schools and thus excludes community-based ECD sites.

How to cite this article: Samuels, M., Taylor, S., Shepherd, D., Van der Berg, S., Jacob, C., Deliwe, C.N. *et al.*, 2015, 'Reflecting on an impact evaluation of the Grade R programme: Method, results and policy responses', *African Evaluation Journal* 3(1), Art. #139, 10 pages. <http://dx.doi.org/10.4102/aej.v3i1.139>

Copyright: © 2015. The Authors. Licensee: AOSIS OpenJournals. This work is licensed under the Creative Commons Attribution License.

14 April 2014. An improvement plan based on the results of the evaluation and on stakeholder consultation accompanied the management response. The evaluation was submitted through the Cabinet approval process (cluster, Cabinet committee, Cabinet) and was tabled at a Cabinet meeting on 19 March 2014. Parliament received the report in July 2014, at which time it was placed on the DPME website.

The rest of this introduction describes the motivation behind and the expansion of the Grade R programme. The next section reports on the findings of the impact evaluation. Following that, we reflect on the process of implementing the recommendations flowing from the evaluation findings. The final section concludes.

Motivation and policy process behind the Grade R programme

In 1995, White Paper 1 on Education and Training proposed the establishment of a national system of provision of a compulsory reception year as part of the transformation of education and training (Department of Education 1995). This policy direction was affirmed in the 2001 Education White Paper 5 on Early Childhood Education. Its thrust to provide wider access to ECD programmes has been followed consistently since then.

The conditions in South Africa in 1994 as well as the expected benefits of early interventions were well articulated in the Education White Paper 5 on Early Childhood Education:

Approximately 40% of young children in South Africa grow up in conditions of abject poverty and neglect. Children raised in such poor families are most at risk of infant death, low birth-weight, stunted growth, poor adjustment to school, increased repetition and school dropout. This factor makes it even more imperative for the Department of Education to put in place an action plan to address the early learning opportunities of all learners but especially those living in poverty. Timely and appropriate interventions can reverse the effects of early deprivation and maximise the development of potential.

The policy focus on the state provision of early learning opportunities for low socio-economic status children represented a shift from the approach taken by previous governments (prior to 1994), in which ECD of non-white children was largely left to parents and non-governmental organisations. There was, however, state-funded provision for white children in public pre-schools. The 1996 Interim Policy on Early Childhood Development estimated that about 9% of all South African children from birth to six years had access to public or private ECD facilities. The impact of the history of discriminatory provision meant that at that time 1 in 3 white infants had access to ECD services compared with about one in 8 Indian and mixed race children and one in 16 African children (Department of Education 1996).

The first intervention to realise the objectives articulated in Education White Paper 1 was the National Early Childhood Development Pilot Project, which the then Department of

Education launched in 1997. The overall pilot was designed to test the interim ECD policy, particularly related to the reception year (referred to as Grade R). The pilot's main objectives included these:

- Designing and testing innovations in the ECD field related to interim accreditation, interim policy and subsidy systems;
- Promoting outcomes-based education and assessment in ECD in line with the National Qualifications Framework (NQF);
- Building capacity to administer large-scale provision of ECD, particularly at provincial department level, in conjunction with research and training organisations (RTOs);
- Assuring quality community-based efforts in ECD through subsidies and training;
- Ensuring children receive quality reception year education; and
- Researching the most effective means of delivering the reception year.

Under the National ECD Pilot Project, the provinces were provided with pro-rata funds of about R40 million (approximately US\$4 million) to provide additional subsidies to community-based ECD sites in order to contract training organisations to provide training towards the accreditation of practitioners and to fund provincial monitoring activities. A total of 2730 sites and practitioners were selected by the provinces to participate, affecting approximately 66 000 learners. It is worth noting that the evaluation of this pilot project was weak and the impact of the activities on learners was not measured.

A nationwide audit of ECD provisioning conducted in 1998 (Department of Education 2001) highlighted the fact that the main challenge facing the government was to convert these efforts into a government-wide national programme of action on ECD. The audit concluded that access to ECD provision in South Africa was low and unequal. It also suggested that a base of sites existed from which to expand access and develop quality improvements.

The ECD Conditional Grant² was introduced in the 2001 and 2002 financial year with the aim of extending the provision of a reception year programme. This initially targeted 4500 registered community-based ECD sites, affecting 135 000 children over a three-year period. The main outcome of the process was the provision of learning materials, the training of practitioners, and advocacy to provide key messages. A report on the Conditional Grant (KPMG 2005) noted that it was difficult to measure the impact of the grant since a plan for evaluation had not been determined at the start of the grant nor was baseline data available.

Rapid expansion of the programme

In 2002 less than 40% of 5-year-olds in South Africa were attending an educational institution. By 2011, this figure

²'ECD Conditional Grant' refers to a grant from the Department of Finance to the Department of Education that is specifically earmarked for the ECD programme.

had risen to more than 80% and it is still rising. This dramatic increase in educational participation amongst young children may have been driven by a variety of factors affecting the supply of and demand for early education, but was probably largely due to the deliberate roll-out and rapid expansion of the Grade R programme. Between 2001 and 2012 the numbers enrolled in Grade R programmes at ordinary schools increased more than threefold, from 242 000 to 768 000. The largest increases in educational participation amongst 5-year-olds were experienced in the poorer provinces – Mpumalanga (138% increase between 2002 and 2011), North-West (153%) and Northern Cape (263%).

It would now appear that access to a Grade R programme is near-universal. Based on an analysis of household survey data, it is estimated that the proportion of Grade 1 children who have previously attended Grade R is about 95% (including Grade R at schools and other institutions such as community centres). Participation in Grade R may have been encouraged through the growing number of no-fee schools and through the National Schools Nutrition Programme, which provides a daily meal to children in the majority of schools serving low socio-economic status communities.

Prior to the impact evaluation described in this article, 'monitoring and evaluation' activity was largely limited to monitoring and standard forms of reporting. Two factors previously made evaluation of the programme impact difficult. First, relevant outcomes data were not systematically collected. Secondly, the programme was rolled out in a haphazard sequence such that schools and children selected themselves into being part of the Grade R programme. This meant that there was no comparison group of non-beneficiaries who could legitimately be compared to beneficiaries. This situation remains the norm across education programmes and throughout government, where measurement of programme impact on beneficiaries is rarely conducted.

Impact evaluation

The full report on the impact evaluation of the introduction of the Grade R programme is available on the DPME website.

Literature review

The first component of the impact evaluation was to conduct a review of South African and international literature in order to assess the evidence about the benefits of ECD programmes. The major conclusions from the literature review are summarised in this section.

The first few years of a child's life lay a foundation for cognitive functioning, behavioural, social and self-regulatory capacities, and physical health – early determinants of development that reinforce each other (Richter *et al.* 2012). Returns on investment are greatest for the young as they have a longer horizon over which to recover investments and because 'skill begets skill' (Heckman 2007). Early investment

in disadvantaged young children could reduce inequality and raise productivity (Heckman & Masterov 2007).

Despite strong empirical evidence on the benefits of early interventions in developed countries (Barnett & Ackerman 2006; Belfield 2004; Karoly, Kilburn & Cannon 2005; Magnuson, Ruhm & Waldfogel 2007), much less is known about developing countries (Dawes, Biersteker & Irvine 2008). One reason for this, as highlighted by Alderman (2011), is the difficulty in distinguishing the programme impact from the impact of self-selection: better subsequent school achievement for those who attended preschool often merely reflects the fact that children from families who value education perform better at school.

In Argentina, one year of pre-primary education increased third grade test marks in standardised mathematics and Spanish tests by 23% of a standard deviation (Berlinski, Galiani & Gertler 2009). In Uruguay, children who had attended pre-school had by age 16 completed one more year of education than their siblings who had not, and were 30% less likely to have dropped out (Berlinski, Galiani & Manacorda 2008), whilst Aguilar and Tansini (2012) found that pre-school attendance was a major factor explaining school performance. Studies on the impact of ECD on child outcomes in South Africa cover mainly health benefits. The Sobambisana programmes' impact on children's readiness for Grade R was mixed, with factors beyond the programmes' control often tempering results (Dawes, Biersteker & Hendricks 2011).

The developmental trajectory of most children is well established at school entry: schooling simply reinforces emerging developmental trends and usually widens gaps (Feinstein 2003). Emergent literacy in pre-school (e.g. ability to manipulate phonemes and to recognise letters and letter sounds) and emergent numeracy skills (counting, number knowledge, estimation, number pattern facility) predict later reading achievement and mathematical competence (Duncan *et al.* 2007; Welsh *et al.* 2010) (although these relationships may not necessarily be *causal*). The key question is how much educational intervention *before* primary school can reduce gaps. The opportunity for language learning is greatest *before* children enter school. A South African study found that language delays remained stable between Grades R and 3, suggesting that education was not powerful enough to overcome an entrenched problem (Klop 2005, in Biersteker 2010).

Most low socio-economic status South African children are inadequately prepared for school and experience 'special needs' when entering school (Naudé, Pretorius & Viljoen 2003). However, the literature suggests that simply providing Grade R of any sort is not the answer: there is international evidence that poor-quality ECD may lead to worrying outcomes, including negative, aggressive behaviour, poor language development (Currie 2001), and greater developmental risks (Leseman 2002). De Witt, Lessing and Lenyai (2006) found that 65% of Grade R learners do not meet the minimum criteria for early literacy

development and will enter Grade 1 without the skills or concepts to master reading. It is strongly argued by some that Grade R should be aligned with ECD pedagogical practice and not become a 'watered-down' Grade 1 (Excell & Linington 2011). In the United States, the benefits of Head Start fade more quickly for black children because they attend poorer-quality schools (Currie 2001), indicating that the impact of ECD may depend, in part, on the quality of the school system (Alderman 2011). Many South African children arrive in formal school with their developmental potential considerably compromised and may consequently not be able to benefit much from education in poor-quality schools (SAIDE 2010).

Empirical methodology

The big challenge in measuring the impact of the Grade R programme was to identify a credible estimate of the counterfactual, that is, what outcomes would have been obtained by children who participated in Grade R had they not participated in Grade R? The cleanest method for estimating programme impact that is sometimes used in impact evaluations is to conduct what is known as a randomised controlled trial. In this method, a lottery is used to randomly assign individuals or groups to participate in a particular programme and others to represent a comparison or 'control' group. Borrowing from the nomenclature of medical trials, outcomes in the 'treatment group' are then compared to outcomes in the 'control group'. Since assignment to treatment and control groups is random, there is no reason to expect any systematic differences between the two groups, and consequently any observed differences in outcomes after the implementation of the treatment can be attributed to the treatment.

This evaluation was limited by the fact that the Grade R programme was not implemented with an evaluation in mind. In other words, assignment to the Grade R programme was not random. As a result, children who participated in Grade R cannot simply be compared with children who did not attend Grade R, as these two groups are likely to differ systematically. The research team therefore had to make use of existing datasets and estimate the impact of Grade R attendance based on comparison groups of children who did not attend Grade R that were *as similar as possible* to those who did attend.

The dataset used in the analysis was obtained by merging two data sources to the EMIS master list of primary schools in South Africa.³ The first of these is the SNAP survey that provides information annually on the numbers of learners registered for each grade (including Grade R) in all South African schools; the second, the Annual National Assessments (ANA) of 2011 and 2012, provides test performance in mathematics and language for Grades 1–6. The full dataset comprises 18 102 schools. The EMIS data provides further information on the location of the school,

3. Some 76.4% of schools providing primary school education (Grades 1–7) are primary schools, with the remaining 23.6% being a combination of combined and intermediary schools.

TABLE 1: Number and/or proportion of schools with captured performance by grade.

| Grade | 2011 | | 2012 | |
|---------|--------|------|--------|------|
| | Number | % | Number | % |
| Grade 1 | 7465 | 41.2 | 14 769 | 81.6 |
| Grade 2 | 7150 | 39.5 | 14 865 | 82.1 |
| Grade 3 | 6933 | 38.3 | 14 574 | 80.5 |
| Grade 4 | 7049 | 38.9 | 14 487 | 80.0 |
| Grade 5 | 7042 | 38.9 | 14 089 | 77.8 |
| Grade 6 | 5842 | 32.3 | 15 178 | 83.9 |

Source: Own calculations from ANA 2011 and 2012 data

TABLE 2: Number/proportion of schools with captured performance by province.

| Province | Total number of schools | 2011 (%) | 2012 (%) |
|---------------|-------------------------|----------|----------|
| Western Cape | 1169 | 78.5 | 88.9 |
| Northern Cape | 407 | 57.0 | 80.1 |
| Free State | 992 | 40.0 | 53.4 |
| Eastern Cape | 4772 | 4.3 | 52.1 |
| KwaZulu-Natal | 4222 | 26.3 | 49.6 |
| Mpumalanga | 1323 | 8.0 | 53.1 |
| Limpopo | 2605 | 6.3 | 47.2 |
| Gauteng | 1551 | 27.7 | 50.4 |
| North-West | 1061 | 14.5 | 65.2 |

Source: Own calculations from ANA 2011 and 2012 data

sector of the school, school quintile and school fees charged by the school.⁴ The combined dataset is sufficiently large to permit precise measurement; thus it should facilitate the measurement of even a relatively small impact, should it occur.

Table 1 indicates the number and proportion of schools for which ANA test performance was captured in various grades. In 2011, roughly 33% – 40% of all grade classes were tested and their performance in both tests was captured.⁵ In 2012, the capturing rate had more than doubled. Table 2 shows clear differences in the ANA collection across provinces. Data capturing was particularly weak within the Eastern Cape, Limpopo and Mpumalanga provinces. Differences in the capturing of learner performance in the ANA tests were also evident across the official school poverty quintile classification. Amongst quintile 1, 2 and 3 schools, which represent roughly the poorest 70% of all schools nationally, test results for 2011 were completely missing in about half of schools. This compares to approximately 30% of quintile 4 and 5 schools that did not have test results. In 2012, there was a marked improvement in the proportion of schools with captured data, particularly amongst the lower quintile schools (Table 3).⁶

The SNAP survey indicates the number of children who were enrolled in Grade R in each year. However, there is no way

4. 'Location of school' refers to an urban/rural distinction; 'sector' refers to the public/independent school distinction; and 'school quintile' refers to the official poverty classification of schools into five categories of socio-economic status. The majority of schools (in quintiles 1–3) are non-fee-paying schools, but school fee data is collected in EMIS for schools that do charge fees, which vary widely in the amount charged.

5. It is assumed that all 18 102 schools could potentially have tested all six grades in the ANA, although this is unlikely to have been the case.

6. The final dataset used for this analysis is therefore the population of schools with captured ANA data. It is possible that the impact of Grade R would be different among schools without captured ANA data, as this may be a select sub-sample.

TABLE 3: Proportion of schools tested and data captured by grade in 2011 and 2012.

| School quintile and year | Number of grades tested | | | | | | |
|--------------------------|-------------------------|-----|-----|-----|------|------|------|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| 2011 | | | | | | | |
| Quintile 1 | 48.3 | 4.0 | 5.3 | 6.5 | 6.8 | 11.5 | 17.6 |
| Quintile 2 | 49.5 | 4.7 | 5.6 | 6.9 | 7.0 | 9.4 | 17.0 |
| Quintile 3 | 51.3 | 4.9 | 5.5 | 8.0 | 5.4 | 8.1 | 16.9 |
| Quintile 4 | 28.9 | 1.3 | 3.9 | 4.6 | 3.7 | 7.5 | 50.1 |
| Quintile 5 | 26.6 | 1.4 | 3.8 | 2.7 | 4.4 | 5.4 | 55.7 |
| 2012 | | | | | | | |
| Quintile 1 | 0.0 | 4.6 | 8.4 | 4.7 | 7.9 | 18.8 | 55.6 |
| Quintile 2 | 0.0 | 5.9 | 9.0 | 4.0 | 9.1 | 20.3 | 51.7 |
| Quintile 3 | 0.0 | 5.0 | 9.4 | 5.7 | 10.4 | 15.9 | 53.6 |
| Quintile 4 | 0.0 | 2.9 | 5.7 | 5.1 | 4.7 | 14.3 | 67.3 |
| Quintile 5 | 0.0 | 1.4 | 4.8 | 4.5 | 3.8 | 9.5 | 76.0 |

Source: Own calculations from ANA 2011 and 2012 data

of knowing whether an individual learner identified in the ANA dataset had attended Grade R. Therefore, the best one can do is to derive a proxy measure of ‘treatment’ using the number of Grade R enrolments in the year that a specific learner would have attended the grade, if the learner had not repeated a grade since then. Treatment is calculated as:

$$R_{git} = \frac{\# \text{ of learners in Grade R}_{(t-g)_i}}{\# \text{ of learners in Grade 2}_{(t-g+2)_i}}, \quad t = 2011, 2012 \quad [\text{Eqn 1}]$$

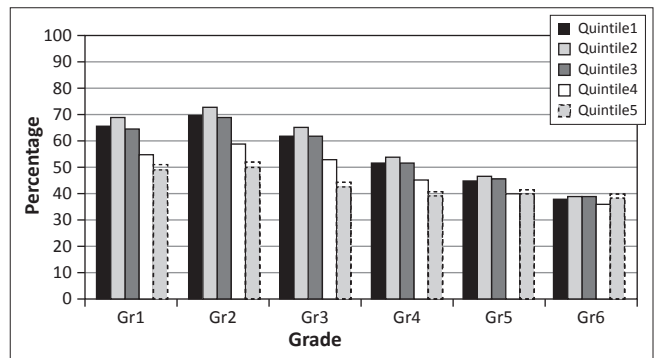
Where:

g is the current grade of the learners and i is the school, $i = 1, 2, \dots, N$.⁷ [Eqn 2]

A number of caveats need to be mentioned with regard to the derivation of the treatment variable. First, in a small number of instances $R_{git} > 1$ which may signal that a school provides Grade R to a wider catchment of learners than actually remain in the school beyond Grade R. In the analysis, R_{git} was top censored to a maximum of 1. A second complicating issue is that some learners may have received Grade R at another educational facility other than the school they were attending at the time of writing the ANA tests. In such cases, R_{git} will underestimate the extent of treatment. Finally, where data for the number of learners in Grade R is missing, it was assumed that no treatment occurred.

Figure 1 indicates average treatment by school quintile and grade in 2012. A comparison of treatment in later grades reflects the expansion of Grade R provision over time. Amongst quintile 1 and 2 schools, for example, close to 70% of Grade 1 learners had attended Grade R, according to this measure, compared with less than 40% of Grade 6 learners. It is interesting to note that the provision of Grade R is lowest amongst quintile 5 schools. This may be influenced by the use of private institutions offering these services to learners from wealthier socio-economic backgrounds. Treatment may therefore be underestimated in the case of learners attending quintile 5 schools.

7. Grade 2 has been used as the denominator in equation (1) rather than Grade 1 due to the high levels of repetition in Grade 1, which inflates the numbers enrolled relative to the underlying cohort size. Using Grade 1 enrolments would therefore cause an underestimate of treatment if this grade were used as the denominator.



Source: Own calculations from SNAP 2005–2012 and EMIS master list

FIGURE 1: Average treatment by school quintile and grade.

The outcome of interest is the mean test score obtained by a particular grade in a school in a particular year, Y_{git} . However, in order to enable comparison across grades, test scores were converted to have a mean of 0 and a standard deviation of 1. The impact of Grade R could then be estimated using a regression analysis, where the standardised test score of an individual is the outcome measure, and explanatory variables include controls for the year of testing (2011 or 2012), the grade of the student, various school characteristics, and the treatment variable, which is the focus of the analysis. The size and significance of the estimated coefficient on the treatment variable represents the impact of having attended Grade R.

The first type of regression model to be estimated is an ordinary least squares (OLS) regression. However, the estimated treatment effect may be biased if any unobservable (or unmeasured) school quality characteristics are correlated with both test scores and the Grade R treatment variable. This could occur if schools providing Grade R self-select into ‘treatment’ based on unobserved dimensions of school quality. For example, it is possible that better-managed schools would have been able to introduce Grade R earlier, whilst such schools may also benefit in terms of their performance. Conversely, attempts by the authorities to expand Grade R rapidly in low socio-economic status schools may have increased treatment in those schools where performance lags. This all means that it is not valid to estimate the impact of Grade R based on a comparison of schools that introduced Grade R early on with schools that introduced it later on or did not introduce Grade R at all.

Given that test scores are observed for Grades 1–6 in two years, there are potentially 12 observations for each school. This makes it possible to use the variation in treatment across grades within schools to identify the treatment effect, whilst correcting for unobserved school characteristics using a school fixed effects (SFE) model.⁸ In effect, this method

8. This amounts to including school-specific dummy variables as explanatory variables in the regression equation, thereby yielding a unique intercept for each school that captures the full effect of school quality or other unobservable school-level factors. Assuming that learner and teacher characteristics of grades within a school are uncorrelated with the Grade R treatment variable, one may posit that controlling for school quality through SFE approximates the impact of our treatment of interest fairly well. Strictly speaking, however, the coefficient on treatment should not be interpreted as truly causal, since assignment to varying levels of Grade R treatment was not random as in an experiment. By controlling for school quality, however, this SFE approach succeeds in eliminating one major source of potential bias. A similar

TABLE 4: OLS and SFE regression results.

| Dependent variable | OLS | | | SFE | | |
|--|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| | 2011 sample | 2012 sample | Pooled sample | 2011 sample | 2012 sample | Pooled sample |
| Standardised mathematics test score | | | | | | |
| Treatment (R_{git}) | 0.199*** (0.020) | 0.145*** (0.015) | 0.159*** (0.012) | 0.074*** (0.018) | 0.025** (0.013) | 0.053*** (0.011) |
| School/province controls | Yes | Yes | Yes | No | No | No |
| Grade fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| School fixed effects | No | No | No | Yes | Yes | Yes |
| Observations | 14 954 | 32 740 | 47 694 | 41 451 | 87 959 | 129 410 |
| R-squared | 0.267 | 0.223 | 0.230 | 0.001 | 0.000 | 0.001 |
| Standardised home/first language test score | | | | | | |
| Treatment (R_{git}) | 0.153*** (0.020) | 0.165*** (0.014) | 0.151*** (0.012) | 0.060*** (0.018) | 0.102*** (0.012) | 0.093*** (0.011) |
| School/province controls | Yes | Yes | Yes | No | No | No |
| Grade fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| School fixed effects | No | No | No | Yes | Yes | Yes |
| Observations | 14 957 | 32 739 | 47 696 | 41 461 | 87 958 | 129 419 |
| R-squared | 0.306 | 0.338 | 0.315 | 0.001 | 0.001 | 0.001 |

Robust clustered standard errors in parentheses.

*, significant at 10% level; **, significant at 5% level; ***, significant at 1% level

measures the correlation between test performance and the Grade R treatment variable within each school separately, and then calculates the average of all within-school correlations. This way, any differences between schools do not affect the estimate of the impact of Grade R.

Results

We begin the analysis with estimates of several OLS regression and SFE models (Table 4). Given that the primary interest of this study is in the impact of Grade R provision, only the regression coefficient on the treatment variable is shown. As a reminder, the dependent variable in all models is the standardised test score. Treatment is coded as a ratio that lies between 0 (no treatment) and 1 (full treatment). This implies that the estimated coefficient on the treatment variable indicates the proportion of a standard deviation change in average test score associated with increasing treatment from zero to full treatment (100% of the cohort having undergone Grade R).

A pooled (2011 and 2012 data combined) OLS model (column 3) indicates a positive and statistically significant coefficient on treatment of approximately 15% of a standard deviation for both mathematics and home language.⁹ When we correct for confounding factors by including school-level fixed effects, the estimated treatment effect is substantially reduced, yet remains statistically significant (reported in columns 4–6). This indicates that much of the association between Grade R attendance and test scores seen in columns 1–3 is actually attributable to unobserved aspects of school quality within the schools that introduced Grade R earlier.

The treatment effect on mathematics score is estimated to be three times greater in 2011 than in 2012. It has already been shown that quintile 4 and 5 schools were over-represented in the 2011 and 2012 samples, and we know the 2011 sample of schools to be on average better performing. It is therefore suspected that using the pooled sample may distort the treatment effect. For consistency, the analysis from this point focuses primarily on results based on the 2012 sample.

Using only the 2012 sample of schools, treatment is estimated to have an impact of 2.5% and 10.2% of a standard deviation respectively on mathematics and language test scores.¹⁰ Filmer, Hasan and Pritchett (2006) have described 40% of a standard deviation as being roughly equal to one grade level in school. Therefore, the estimates here indicate an improvement in average performance equivalent to somewhere between 12 days (for mathematics) and 50 days (for language) of a year's learning, respectively, for having all learners enrol in Grade R. Note that this is an average effect over all grades.

In order to capture possible differences in school functioning within school quintiles, the sample was sub-divided into four groups: quintile 1–4 schools in weaker performing provinces; quintile 5 schools in weaker performing provinces; quintile 1–4 schools in top performing provinces; and quintile 5 schools in top performing provinces.¹¹ This sub-division was based on the premise that the top performing provinces may face fewer constraints with regard to the functioning of school-based programmes. The results of fixed effects regression based on these samples are shown in Table 5.

Attending Grade R is estimated to have a positive and statistically significant effect across all four sub-samples.

⁹SFE method is employed by Taylor and Coetzee (2013) to evaluate the impact of language of instruction in Grades 1–3 on subsequent learning outcomes in South Africa.

⁹The unavailability of data on school fees in the OLS considerably reduces the sample. In the SFE models it is unnecessary to include any school-level characteristics, such as fees.

¹⁰In both mathematics and language one standard deviation is approximately 20 percentage points in the ANA tests, with some variation, depending on the grade.

¹¹The top performing provinces here identified are Gauteng, Northern Cape and Western Cape, with the remaining six provinces falling in the weaker performing group.

TABLE 5: Effect of treatment, by school wealth quintile and province.

| Dependent variable | Weak performing provinces | | Top performing provinces | |
|--|---------------------------|---------------------|--------------------------|---------------------|
| | Quintiles 1–4 | Quintile 5 | Quintiles 1–4 | Quintile 5 |
| Standardised mathematics test score | | | | |
| Treatment (R_{gr}) | 0.018*** (0.007) | 0.096*** (0.045) | 0.104*** (0.030) | 0.160*** (0.051) |
| Grade fixed effects | Yes | Yes | Yes | Yes |
| School fixed effects | Yes | Yes | Yes | Yes |
| Observations | 54 095 | 3219 | 10 786 | 3179 |
| R-squared | 0.002 | 0.030 | 0.009 | 0.239 |
| Standardised home/first language test score | | | | |
| Treatment (R_{gr}) | 0.030*** (0.007) | 0.133*** (0.049) | 0.041 (0.032) | 0.137*** (0.052) |
| Grade fixed effects | Yes | Yes | Yes | Yes |
| School fixed effects | Yes | Yes | Yes | Yes |
| Observations | 54 095 | 3219 | 10 786 | 3179 |
| R-squared | 0.023 | 0.275 | 0.134 | 0.679 |

Robust clustered standard errors in parentheses. 2012 ANA sample only.

*, significant at 10% level; **, significant at 5% level; ***, significant at 1% level

However, there are noticeable differences in the magnitude of the effect. Treatment is estimated to increase average mathematics performance by 1.8% of a standard deviation in the case of poorer schools in weak performing provinces compared to 9.6% of a standard deviation for quintile 5 schools in the same provinces. The latter effect is numerically equivalent to the impact of Grade R in poorer schools in the top performing provinces, suggesting that programmes such as Grade R provision provide greater benefits when implemented within a well-functioning education system, even in the poorer schools in such provinces. The wealthiest schools in the top performing provinces have the largest positive impact of treatment in mathematics performance at 16% of a standard deviation. Similar results are found for home language in that the effect of treatment is smaller for quintile 1–4 schools (3% – 4% of a standard deviation) compared to quintile 5 schools (13% of a standard deviation). However, unlike mathematics performance, there do not appear to be any statistically significant differences in the effect of treatment across the two provincial groupings within the same school wealth quintiles.

In summary, there was an overall positive impact of Grade R on later learning outcomes in both language and mathematics, though the size of the effects was small relative to what one might have hoped to see. In some schools Grade R has contributed towards better learning, but in other schools it has not. These findings confirm that in Grade R, as is the case throughout the school system, there are significant challenges to ensuring instructional quality. This is truest of the parts of the school system serving poor learners, where the estimated impact of Grade R was almost negligible.

One further limitation of this data analysis should be noted. Whilst the Grade R programme is intended to have multiple benefits, including physical, mental, emotional, social and moral development (according to the 1995 White Paper), the only measurable outcomes for this study were mathematics and language performance as measured in the ANA.

Recommendations

One limitation of this evaluation was that it was not able to identify (in a quantitative way) reasons *why* Grade R did or did not have an effect in particular schools or groups of schools. The data and method allowed for an estimation of the impact of Grade R attendance on later learning outcomes but was unable to unpack the intermediary causal mechanisms. This means that the recommendations following from the evaluation are fairly generic and do not follow directly from the quantitative analysis.

Whilst it would have been preferable for the Grade R impact evaluation to have included a qualitative component (for instance, a survey focusing on implementation) to shed light on the reasons behind the observed effect sizes, the fact that we now have a sense of the magnitude of the impact is still immensely valuable. Previous and future studies that describe the challenges and successes in implementing the Grade R programme can now be interpreted within the context of knowing the overall magnitude of programme impact.

Therefore, some recommendations for strengthening the Grade R programme are made, taking into consideration current policy questions about Grade R and on the basis of other studies, including a public expenditure tracking study undertaken in 2011 by the same research team from the University of Stellenbosch. These recommendations are to a large extent reflected in the Grade R improvement plan that was developed in response to the evaluation.

The first recommendation is that an interim Grade R policy should be developed for submission to the Cabinet. The policy should provide clarity on, amongst other aspects: (1) age of admission/school readiness; (2) role of community-based sites; (3) funding; (4) employment of Grade R teachers; (5) infrastructure, and (6) learners with disabilities.

Establishing a clear picture on how much government spends on the Grade R programme is difficult due to inconsistencies and inaccuracies in the way provincial

education departments record spending. In some provinces, the reported per pupil spending on Grade R is very low, probably because of cross-subsidisation of Grade R from other programmes or anomalies in how Grade R spending is categorised. Provincial financial record-keeping should be attended to urgently and then regularly analysed so as to inform planning.

An audit of pre-service training opportunities for Grade R practitioners should be conducted (of, for instance, the Grade R diploma offered at FET colleges). This should contribute to an understanding of the numbers graduating from such programmes and of the appropriateness of their content.

Since there are already many Grade R practitioners in schools, opportunities for in-service training need to be increased. These should be focused on providing teachers with practical strategies for supporting early learning and opportunities to observe best-practice teaching. Ideally, this needs to be supported with continuous on-site mentoring. However, it may not be feasible to provide good-quality on-site support at full scale. This is not to say it should not be considered in a limited section of high-priority schools, such as quintile 1 schools in certain provinces. Less costly teacher support innovations also need to be developed, such as resource packs with practical strategies to apply.

Culturally relevant storybooks in all South African languages should be made more widely available to parents and/or caregivers, in particular through community libraries and in Grade R classrooms.

A high-quality school readiness test should be developed or identified and this should be provided to Grade R practitioners to use as a tool in assessing the development of their children. An emphasis on the use of such a tool will help to raise the awareness amongst schools, parents and practitioners that certain clear developmental outcomes must be obtained during Grade R – that it is not sufficient for children simply to attend a type of crèche.

There are other policy questions to which this evaluation does not provide clear answers. For instance, there is some debate about whether to prioritise the expansion of a 'pre-Grade R' year. The finding of low impact of Grade R may point to the need to improve quality before expanding access to even younger children. However, in some schools there are already large numbers of under-aged children attending Grade R and they often attend for two years. Separating younger and older children into two classes with separate and appropriate curriculum may actually help improve the effectiveness of Grade R itself. To some degree, therefore, this impact evaluation raises further questions.

Implementation of findings

According to the processes prescribed in the NEP, once the impact evaluation report was finalised, a team of national and

provincial officials, as well as several external experts, met to compile an improvement plan for the Grade R programme. The improvement plan, based on the recommendations made in the report, includes the following activities:

- development of an interim Grade R policy;
- development of a human resource strategy;
- support for curriculum implementation, including the provision of materials;
- development of an integrated monitoring and evaluation system.

The improvement plan has been signed by the director-general of the DBE. Although at the time of writing this paper it has only been about nine months since the development of the improvement plan, it is already fair to say that progress in implementation has been slow.

The improvement plan recommended as a starting point that a task team comprising various branches within the department be set up to drive the development of an interim policy and human resource strategy. This has not been instituted yet and has therefore held back the delivery of the improvement plan. The reality is that little progress tends to occur until senior officials drive processes, but their attention is divided between a range of priorities. The last nine months have also included a national election, substantial restructuring within the DBE and the tenure of two acting director-generals. In this context it is perhaps understandable that progress has been slow.

Another institutional reality of policy formulation is that numerous political groupings and stakeholders are simultaneously pushing different agendas. This impact evaluation and its recommendations is only one such process. There is also the National Development Plan, which, for example, recommends a second year of pre-schooling. The ruling party has its own processes for identifying policy direction. Teacher unions also have certain agendas. As a result, the recommendations flowing from this evaluation are only one consideration amongst many in the policy formulation process. To illustrate, one recommendation ensuing from this evaluation is that support for Grade R practitioners should focus on practical strategies for supporting early learning and opportunities to observe good teaching. There may, however, be pressure through other processes to focus on upgrading the paper qualifications of Grade R practitioners and increase their remuneration. The matter is no doubt complex and an innovative strategy will have to balance the needs of the children against those of the adults working with them.

Arguably, the most significant effect of conducting evaluations within government, such as this one, is to foster a culture in which the focus of policymakers and programme managers gradually shifts towards programme outcomes rather than only programme inputs. In the DBE, there have been a few impact evaluations over the past couple of years (Department of Basic Education 2013; Fleisch *et al.* forthcoming; Taylor &

Coetzee 2013). These have demonstrated that well-designed programmes often have low or even negligible impacts on learning outcomes, and that programme implementation cannot be taken as a guarantee of programme impact. The role of the DPME and the NEP in initiating this evaluation was critical to ensuring that an evaluation with measurement of programme impact took place, and was made available publicly. This no doubt strengthens accountability and knowledge within the basic education sector.

One should not be naive about the incentives facing the government when conducting evaluations. Often, as in the case of the Grade R impact evaluation, the results can point to significant problems and low impact. In an environment where the media are likely to pick up on this and create negative press for the implementing department, this creates an incentive for government officials to resent an evaluation rather than embrace it so as to learn from it. The DPME will need to find ways to assist partnering departments in communicating findings to the public and in ensuring that the process is constructive.

Conclusion

The major success of the Grade R programme has been how rapidly it was expanded since 2001, especially in the poorer parts of the country. As with most government programmes, Grade R was not rolled out with impact evaluation in mind, which could have allowed clear intervention and comparison groups to be identified. Therefore, when the DPME and the DBE placed the Grade R programme on the NEP, the research methodology was inevitably complicated statistically and reliant on the data that was available. Nevertheless, the research conducted by independent academics at the University of Stellenbosch has provided what can be interpreted as a fairly reliable indication of the causal impact of the Grade R programme on later learning outcomes.

Attending Grade R was associated with better language and mathematics performance during primary school. However, the impact was fairly small and nearly negligible in low socio-economic status schools located in poorer provinces. This is unfortunate since the Grade R programme was intended to reduce the educational disadvantage faced by low socio-economic status children.

The finding of low-quality delivery of Grade R in low socio-economic status schools in poorer provinces is consistent with the systemic challenges observed in primary and secondary schools in these contexts. Various researchers describe the South African school system as consisting of two sub-systems (Fleisch 2008; Spaull 2013). There is a fairly well-performing section of the school system, consisting of historically white and Indian schools and serving predominantly middle class children. Then there is a majority group of historically disadvantaged schools serving low socio-economic status communities whose learners perform below minimum standards on

average and which display teacher and organisational characteristics that leave much to be desired – for instance; high levels of teacher absenteeism, low levels of teacher subject knowledge and low time-on-task. In these weakly functioning schools the impacts of resources and interventions are often low, since other binding constraints preclude their effectiveness. For example, Van der Berg (2008) argues that additional resources matter conditionally upon school management.

Despite the limitations of the data and methodology employed in this impact evaluation, it represents a major advance on what was previously possible. The impact evaluation has demonstrated the value of administrative data, even though such data are not perfectly clean. The ANA and the National Senior Certificate data provide education outcomes data for the population of schools and students. All that is needed is for programme delivery to be implemented in a sequence that allows for identification of the beneficiaries and a valid control group. More impact evaluations should therefore be possible in future. In the absence of random assignment to programmes, the use of SFE modelling can to some extent facilitate the estimation of programme impact.

In order to improve the quality of the programme, steps must be taken to support Grade R practitioners with practical training, to improve the provision of support materials and to help practitioners to monitor the school readiness of learners. Whilst the implementation of these recommendations may not be a smooth linear process flowing from this particular evaluation, the process of evaluating programmes such as Grade R initiated and conducted within the government is a potentially valuable contribution towards improving service delivery.

Acknowledgements

The authors would like to thank the two anonymous referees for their helpful comments and recommendations for strengthening the paper. We are also grateful to the DPME for funding the impact evaluation of the implementation of Grade R, and to the DBE for granting access to the data and for co-operation on the project.

Competing interests

The authors declare that they have no financial or personal relationship(s) that may have inappropriately influenced them in writing this article.

Authors' contributions

Marie-Louise Samuels, lead author, Grade R programme manager. Stephen Taylor, contributing author, lead co-ordinator of the paper. Debra Shepherd, contributing author, main data analyst.

Servaas van der Berg, contributing author, principal investigator on the impact evaluation itself.

Christel Jacob, contributing author, including around development of NES and process of the evaluation. Carol Nuga Deliwe, contributing author, overseeing the evaluation from DBE. Thabo Mabogoane, contributing author, steering committee and paper coordination.

References

- Aguilar, R. & Tansini, R., 2012, 'Joint analysis of preschool attendance and school performance in the short and long-run', *International Journal of Educational Development* 32(2), 224–231. <http://dx.doi.org/10.1016/j.ijedudev.2011.03.001>
- Alderman, H. (ed.), 2011, *No small matter: The impact of poverty, shocks, and human capital investments in early childhood development*, World Bank, Washington.
- Barnett, W.S. & Ackerman, D.J., 2006, 'Costs, benefits, and long-term effects of early care and education programs: Recommendations and cautions for community developers', *Community Development* 37(2), 86–100. <http://dx.doi.org/10.1080/15575330609490209>
- Belfield, C.R., 2004, *Early childhood education: How important are the cost-savings to the school system?* Center for Early Care and Education, viewed 11 August 2015, from
- Berlinski, S., Galiani, S. & Gertler, P., 2009, 'The effect of pre-primary education on primary school performance', *Journal of Public Economics* 93(1–2), 219–234. <http://dx.doi.org/10.1016/j.jpubeco.2008.09.002>
- Berlinski, S., Galiani, S. & Manacorda, M., 2008, 'Giving children a better start: Preschool attendance and school-age profiles', *Journal of Public Economics* 92(5–6), 1416–1440. <http://dx.doi.org/10.1016/j.jpubeco.2007.10.007>
- Biersteker, L., 2010, 'Scaling up early childhood development in South Africa: Introducing a reception year (Grade R) for children aged five years as the first year of schooling', Wolfensohn Center for Development, Working Paper 17, April 2010.
- Currie, J., 2001, 'Early childhood education programs', *Journal of Economic Perspectives* 15(2), 213–238. <http://dx.doi.org/10.1257/jep.15.2.213>
- Dawes, A., Biersteker, L. & Hendricks, L., 2011, *Towards integrated early childhood development: An evaluation of the Sobambisana initiative*, Ilifa Labantwana, Cape Town.
- Dawes, A., Biersteker, L. & Irvine, M., 2008, *Scaling up early childhood development (ECD) (0–4 years) in South Africa. What makes a difference to child outcomes in the period 0–4? Inputs for Quality ECD Interventions*. Department of Social Development, Pretoria.
- Department of Basic Education, 2013, *Developing and evaluating the first phase of the Grade 12 Mind the Gap study guide series*, Department of Social Development, Pretoria.
- Department of Education, 1995, *White Paper on Education and Training, Government Notice 196 of 1995, 15 March 1995*, Government Printer, Pretoria.
- Department of Education, 1996, *Interim policy on early childhood education*, Government Printer, Pretoria.
- Department of Education, 2001, *White Paper on Early Childhood Education, Government Notice 5 of 2001, 30 May 2001*, Government Printer, Pretoria.
- De Witt, M.W., Lessing, A.C. & Lenyai, E.M., 2006, 'An investigation into early literacy of preschool learners', Unpublished article.
- Duncan, G.J., Dowsett, C.J., Claessens, A., Magnuson, K.A., Huston, A.C., Klebanov, P. et al., 2007, 'School readiness and later achievement', *Developmental Psychology* 43(6), 1428–1446. PMID: 18020822, <http://dx.doi.org/10.1037/0012-1649.43.6.1428>
- Excell, L. & Lington, V., 2011, 'Taking the debate into action: Does the current Grade R practice in South Africa meet quality requirements?', *SA-eDUC Journal* 8(2), 3–12.
- Feinstein, L., 2003, 'Inequality in the early cognitive development of British children in the 1970 cohort', *Economics* 70, 73–97. <http://dx.doi.org/10.1111/1468-0335.t01-1-00272>
- Filmer, D., Hasan, A. & Pritchett, L., 2006, 'A millennium learning goal: Measuring real progress in education', Center for Global Development, Working Paper 97.
- Fleisch, B., 2008, *Primary education in crisis: Why South African schoolchildren underachieve in reading and mathematics*, Juta, Cape Town.
- Fleisch, B., Taylor, S., Schöer, V. & Mabogoane, T., forthcoming, 'The impact of an 11-week teacher support programme on pupil reading proficiency', Unpublished manuscript.
- Goldman, I., Mathe, J.E., Jacob, C., Hercules, A., Amisi, M., Buthelezi, T. et al., 2015, 'Developing South Africa's national evaluation policy and system: First lessons learned', *African Evaluation Journal* 3(1), 1–9.
- Heckman, J., 2007, 'The economics, technology, and neuroscience of human capability formation', NBER Working Paper 13195.
- Heckman, J. & Masterov, D.V., 2007, 'The productivity argument for investing in young children', Institute for the Study of Labor, Discussion Paper 2725.
- Karoly, L., Kilburn, R. & Cannon, J., 2005, 'Early childhood interventions: Proven results, future promise', *RAND Labor and Population*, The Rand Corporation, Santa Monica.
- KPMG, 2005, 'Report on the evaluation of conditional grants administered by the Department of Education', Internal report for the government, not publicly available.
- Leseman, P., 2002, 'Early childhood education and care for children from low-income or minority backgrounds', paper presented at the OECD Oslo workshop, 6–7 June 2002.
- Magnuson, K.A., Ruhm, C. & Waldfogel, J., 2007, 'The persistence of preschool effects: Do subsequent classroom experiences matter?', *Early Childhood Research Quarterly* 22(1), 18–38. <http://dx.doi.org/10.1016/j.ecresq.2006.10.002>
- Naudé, H., Pretorius, E. & Viljoen, J., 2003, 'The impact of impoverished language development on preschoolers' readiness-to-learn during the Foundation Phase', *Early Child Development and Care* 173(2–3), 271–291. <http://dx.doi.org/10.1080/0300443030303098>
- Richter, L., Biersteker, L., Burns, J., Desmond, C., Feza, N., Harrison, D. et al. 2012, 'Diagnostic review of early childhood development', Report for Department of Performance Monitoring and Evaluation (DPME) in the Presidency.
- SAIDE (South African Institute for Distance Education), 2010, *Will Grade R really improve the quality of SA education?*, Grade R Research Project, Final Report.
- Spaull, N., 2013, 'Poverty & privilege: Primary school inequality in South Africa', *International Journal of Educational Development* 33, 436–447. <http://dx.doi.org/10.1016/j.ijedudev.2012.09.009>
- Taylor, S. & Coetzee, M., 2013, 'Estimating the impact of language of instruction in South African primary schools: A fixed effects approach', Stellenbosch Economics Working Papers 21/13.
- Van der Berg, S., 2008, 'How effective are poor schools? Poverty and educational outcomes in South Africa', *Studies in Educational Evaluation* 34, 145–154. <http://dx.doi.org/10.1016/j.stueduc.2008.07.005>
- Welsh, J.A., Nix, R.L., Blair, C., Bierman, K.L. & Nelson, K.E., 2010, 'The development of cognitive skills and gains in academic school readiness for children from low-income families', *Journal of Educational Psychology* 102(1), 43–53. PMID: 20411025, <http://dx.doi.org/10.1037/a0016738>