CHAPTER 3
METHODOLOGY

Research approach

The focus of this study was a quantitative analysis of children’s access to education at both the basic and the further education levels, in other words, for children aged 7 to 18 years of age. It aimed to provide a quantitative comparison of access to education across the nine provinces and for both males and females, by using the 2001 Annual School Survey (ASS) data, the 2001 Snap Survey data and 2001 Census data.

Assessing the level of access to schooling in South Africa was done in two ways. Firstly, an overview of enrolment in school both nationally and provincially was undertaken. Then enrolment and participation rates were determined by using the standard indicators that are used to measure capacity and access. These are gross enrolment ratios, net enrolment ratios, age-specific enrolment ratios, apparent intake rate and net intake rate.

Data analysis indicators

Gross enrolment ratio

The gross enrolment ratio (GER) measures enrolment, regardless of age, in a specific level of education as a proportion of the appropriately aged population for the given level of education.

This is calculated by dividing the number of learners enrolled in a given level of education regardless of age by the population of the age group which officially corresponds to the given level of education, and multiplying the result by 100 (Unesco Institute for Statistics, GER). The formula for calculating GER is as follows:

\[ \text{GER}_{h} = \frac{E_{h}}{P_{h,a}} \times 100 \]

where:
GER\textsuperscript{t} = Gross enrolment ratio at level of education \( h \) in school-year \( t \)
E\textsuperscript{t} = Enrolment at the level of education \( h \) in school-year \( t \)
P\textsuperscript{t,\( a \)} = Population in age-group \( a \) which officially corresponds to the level of education \( h \) in school-year \( t \)

The GER indicates the level of participation in education, regardless of whether or not the learners belong to the official age-group. While a GER value of 100% or more indicates that a country has the capacity to accommodate all of its school-age population, it does not indicate what proportion of that school-age population is actually enrolled. A GER of over 100% is usually due to the inclusion of over-age and under-age learners in the system, either as a result of early or late entrance into the education system or as a result of repetition. GER of 100% is, therefore, a ‘necessary but not sufficient’ indicator (Colclough, 1993, p.17) for measuring universal attendance at school.

Consequently, net enrolment rates will also be calculated.

**Net enrolment ratios**

The net enrolment rate (NER) reflects the number of appropriately aged learners in a specified level of education, as a proportion of the corresponding age group in the population. It excludes all learners above and below the official school age who might be enrolled in the education system. The NER reflects the level of participation of learners of the correct age group as well as, to some extent, the level of internal efficiency of the school system.

It is calculated by dividing the number of learners enrolled who are of the official age-group for a given level of education by the population for the same age group, and multiplying the results by 100 (Unesco Institute for Statistics, NER).

The formula for NER is as follows:

\[
\text{NER}\textsuperscript{t} = \frac{E\textsuperscript{t}}{P\textsuperscript{t,\( a \)}} \times 100
\]
where:
\( \text{NER}^t_h = \text{Net enrolment ratio at level of education } h \text{ in school-year } t \)
\( E^t_{h,a} = \text{Enrolment of the population of age-group } a \text{ at a level of education } h \text{ in school-year } t \)
\( P^t_{h,a} = \text{Population in age-group } a \text{ which officially corresponds to level of education } h \text{ in school-year } t \)

A high NER reflects a high degree of participation in education of the official school-age population. The theoretical maximum NER value is 100%. If the NER is below 100%, then the balance reflects the proportion of appropriately aged children not enrolled in the specified level of education. However, these children may be enrolled at other levels of education, and consequently the balance cannot, on its own, be considered as an indicator of the proportion of children not enrolled in education. For this reason age-specific enrolment ratios also need to be calculated to complement the GER and NER indicators.

**Age-specific enrolment ratios**

Age-specific enrolment ratios (ASER) show the percentage of the population of a specific age enrolled in education, irrespective of the level of education they are enrolled in. It illustrates the extent of participation in educational activities of a specific age cohort. (Unesco Institute for Statistics, ASER)

It is calculated by dividing the number of learners of a specific age enrolled in educational institutions at all levels by the population of the same age, and then multiplying the result by 100.

The formula for ASER is as follows:

\[
\text{ASER}^t_a = \frac{E^t_a}{P^t_a} \times 100
\]

where:
\( \text{ASER}^t_a = \text{Age-specific enrolment ratio of an age } a \text{ in school-year } t \)
\( E^t_a = \text{Enrolment of the population of a specific age } a \text{ in school-year } t \)
\( P^t_a = \text{Population in age } a \text{ in school-year } t \)
A high ASER reflects a high degree of participation in education of the population of a particular age. The theoretical maximum ASER value is 100%. If the ASER is below 100%, then the balance reflects the proportion of children of a specific age who are not enrolled in any level of education.

While ASER is an effective indicator for assessing participation of school-aged children in education, it does not give an indication of the grade or the level of education in which learners are enrolled, which would give an indication of the efficiency or quality of the system.

*Apparent intake rate*

The apparent intake rate (AIR) measures the total number of new entrants in the first grade of primary education, regardless of age, expressed as a percentage of the population at the official primary school-entrance age.

This is calculated by dividing the number of all new entrants into grade 1, regardless of age, by the population of official school-entrance age, and multiplying the result by 100 (Unesco Institute for Statistics, AIR). The formula for calculating AIR is as follows:

\[
\text{AIR}_t = \frac{N_t}{P_a^t} \times 100
\]

where:

\(\text{AIR}_t\) = Apparent intake rate in school-year \(t\)

\(N_t\) = Number of new entrants in the first grade of primary education, in school-year \(t\)

\(P_a^t\) = Population of official primary school entrance age \(a\), in school-year \(t\)

The AIR indicates the level of access to primary school regardless of age. An AIR of more than 100% is due to a large number of over- and under-aged children entering primary school for the first time. The figure could be distorted if repeaters in grade 1 are included in the data.
Net intake rates

The net intake rate (NIR) measures the new entrants in the first grade of primary education who are of the official primary school-entrance age as a proportion of the population of the same age.

It is calculated by dividing the number of children of official school-entrance age who enter the first grade of primary education by the population of the same age, and multiplying the result by 100 (Unesco Institute for Statistics, NIR). The formula for calculating NIR is as follows:

\[ \text{NIR}_t = \frac{N_{ta}}{P_{ta}} \times 100 \]

Where:
- \( \text{NIR}_t \) = Net intake rate in school-year \( t \)
- \( N_{ta} \) = Number of new entrants of official primary school-entrance age \( a \) who enter the first grade of primary education, in school-year \( t \)
- \( P_{ta} \) = Population of official primary school entrance age \( a \), in school-year \( t \)

A high NIR indicates a high degree of access to primary education for children of the official primary school entrance age. However, the indicator can be distorted if repeaters in grade 1 are not distinguished from new entrants. The will be the case especially for under-aged learners who repeat grade 1 at the official entrance age.

The data

Three main sources of data were used to investigate the extent to which access has been achieved in South Africa: the Annual School Survey (ASS), the Snap Survey and the national census.

The ASS is conducted by each of the nine provincial education departments in March or April of every year. The survey collects information on the learners enrolled, and staff employed in every primary and secondary school in the country. The most
relevant data in the annual school survey, for the purpose of this study, is the data on enrolment by grade and age by sex. The data collected in 2001 was used and was obtained from the national DoE.

The second source of data was the Snap Survey, which is conducted every year on the 10th day of the school year. It collects basic information on learners and educators – the information collected in this survey is not as detailed as that collected by the ASS. The most relevant for this study was enrolment of learners by grade and sex. The Snap data for 2001 was used and was obtained from the national DoE.

The third source of data was the national census that was carried out in October 2001 by Statistics SA. This data was obtained from the Census 2001 Community Profile Database, which is made available by Statistics SA for public use. The database was obtained by the Education Foundation, an NGO which undertakes research in education, and where I was able to use the database. The most relevant data from Census 2001 for this study is Present School Attendance by geography, age, and province and Population by geography, age and gender.

By using enrolment and population data from the same year, it was intended that this decision would obviate the necessity of having to develop projections, and would provide a more accurate assessment of the actual situation.

Annual school survey

Permission to use the 2001 ASS data was requested from the DoE in July 2004 (see letter in Appendix 2). While waiting for a response from the DoE, preliminary queries were run on 2001 ASS data which had been given by the DoE to the Education Foundation Trust, an education non-governmental organisation, and to which I had access. The queries were run on the data for enrolment of learners by race, enrolment of learners by home language and enrolment of learners by age, in order to assess the number of learners both provincially and nationally for the different data. The race data showed the largest number of learners enrolled in schools (11 334 068), followed by the age data (11 199 479) and then the home language data (11 156 471). These data were then compared to the 2001 Snap Survey data published
by the DoE in its publication *Education Statistics in South Africa at a Glance in 2001* (Department of Education 2003b, pp.12, 13, 27), which showed that there were 11 540 805 learners enrolled at public ordinary and independent schools from grades 1 to 12 and schools for learners with special educational needs (ELSEN) (see table 1).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastern Cape</td>
<td>2 077 235</td>
<td>2 126 684</td>
<td>2 113 354</td>
<td>2 016 831</td>
</tr>
<tr>
<td>Free State</td>
<td>701 042</td>
<td>705 416</td>
<td>698 096</td>
<td>704 042</td>
</tr>
<tr>
<td>Gauteng</td>
<td>1 503 118</td>
<td>1 517 715</td>
<td>1 453 424</td>
<td>1 563 380</td>
</tr>
<tr>
<td>KwaZulu-Natal</td>
<td>2 672 556</td>
<td>2 718 578</td>
<td>2 700 260</td>
<td>2 629 853</td>
</tr>
<tr>
<td>Limpopo</td>
<td>1 807 746</td>
<td>1 839 852</td>
<td>1 814 535</td>
<td>1 717 034</td>
</tr>
<tr>
<td>Mpumalanga</td>
<td>747 878</td>
<td>726 989</td>
<td>705 761</td>
<td>899 786</td>
</tr>
<tr>
<td>North West</td>
<td>554 938</td>
<td>567 300</td>
<td>563 550</td>
<td>894 656</td>
</tr>
<tr>
<td>Northern Cape</td>
<td>188 856</td>
<td>192 988</td>
<td>191 710</td>
<td>199 313</td>
</tr>
<tr>
<td>Western Cape</td>
<td>946 110</td>
<td>938 546</td>
<td>915 781</td>
<td>915 910</td>
</tr>
<tr>
<td>Total</td>
<td>11 199 479</td>
<td>11 334 068</td>
<td>11 156 471</td>
<td>11 540 805</td>
</tr>
</tbody>
</table>

Sources: Department of Education, 2003b; Department of Education, 2004a

In August, Mr Christo Lombaard, an official in the Education Management Information Systems (EMIS) directorate of the DoE, contacted me telephonically and informed me that the DoE would provide the data that I needed. I explained that I had access to the data through the Education Foundation Trust. However, Mr Lombaard requested that I get the data again from the DoE as he was not sure whether or not any changes had been made to the 2001 ASS database after the data had been given to the Education Foundation Trust. During the same telephone conversation I requested the 2001 Snap Survey database so that comparisons could be made between the Snap Survey and the ASS. This request was also made in an e-mail to Mr Lombaard.

In September, I received from the DoE the age-by-grade data by gender and province by school, the number of first-time enrolments in grade 1 data by age, gender and province by school, and the snap survey data for each province for 2001. Queries
were run on the age-by-grade data and the results were identical to those obtained from the data acquired from the Education Foundation Trust.

Cleaning of the data

The data was not cleaned. Problems with the data that reflected, for example, very young learners enrolled in secondary school grades or old learners enrolled in the lower grades, were left unchanged.

Age-grade tables by gender for each province were created. Learners enrolled in the pre-primary phase (pre-schools and grade R) or in post matric were then excluded, so that the tables reflected only those learners enrolled from grades 1-12 and in ELSEN. These tables were then compared with the Snap Survey data for 2001 published by the DoE (Department of Education, 2003b). The comparison showed that, nationally, the DoE’s published statistics (Department of Education, 2003) based on the 2001 Snap Survey had 6% more learners (or 703 886 learners) than there were in the age-by-grade data in the 2001 ASS database (see table 2). While there were differences in all the provinces, the biggest and most serious differences in the data were in the North West and Mpumalanga provinces. In the North West province there were 342 020 fewer learners in the 2001 ASS data than in the Department’s published Snap data (Department of Education, 2003b), a difference of 38%. In Mpumalanga, there was a difference of 166 610 learners (or 18.5%) between the two sets of data.
<table>
<thead>
<tr>
<th>Province</th>
<th>No. of learners in grades 1 to 12 and Elsen from ASS 2001 (excludes pre-primary phase and post matric)</th>
<th>No. of learners in grades 1 to 12 and Elsen from Statistics at a Glance (excludes pre-primary phase and unspecified)</th>
<th>Difference</th>
<th>Percentage difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastern Cape</td>
<td>2 043 725</td>
<td>2 016 831</td>
<td>-26 894</td>
<td>-1.3</td>
</tr>
<tr>
<td>Free State</td>
<td>683 608</td>
<td>704 042</td>
<td>20 434</td>
<td>2.9</td>
</tr>
<tr>
<td>Gauteng</td>
<td>1 477 383</td>
<td>1 563 380</td>
<td>85 997</td>
<td>5.5</td>
</tr>
<tr>
<td>KwaZulu-Natal</td>
<td>2 575 615</td>
<td>2 629 853</td>
<td>54 238</td>
<td>2.1</td>
</tr>
<tr>
<td>Limpopo</td>
<td>1 691 951</td>
<td>1 717 034</td>
<td>25 083</td>
<td>1.5</td>
</tr>
<tr>
<td>Mpumalanga</td>
<td>733 176</td>
<td>899 786</td>
<td>166 610</td>
<td>18.5</td>
</tr>
<tr>
<td>North West</td>
<td>552 636</td>
<td>894 656</td>
<td>342 020</td>
<td>38.2</td>
</tr>
<tr>
<td>Northern Cape</td>
<td>184 333</td>
<td>199 313</td>
<td>14 980</td>
<td>7.5</td>
</tr>
<tr>
<td>Western Cape</td>
<td>894 492</td>
<td>915 910</td>
<td>21 418</td>
<td>2.3</td>
</tr>
<tr>
<td>Total</td>
<td>10 836 919</td>
<td>11 540 805</td>
<td>703 886</td>
<td>6.1</td>
</tr>
</tbody>
</table>


**Filling in the missing data**

The nine provincial Snap Survey databases for 2001, provided by the DoE, were then used to identify missing schools and learners. The Snap Survey does not provide an age breakdown of enrolment. It does, however, give the number of learners by grade and (with the exception of the Free State) by gender.

From the age-by-grade tables that had been created from the 2001 ASS, percentages of the learners by gender and by age in each grade were calculated. The schools that appeared in the 2001 Snap Survey and not the 2001 ASS were then identified, together with the number of learners by gender and grade. The proportions of learners by age, grade and gender that were calculated with the 2001 ASS data were then applied to the enrolment-by-grade data from the Snap 2001 data, in order to get an age breakdown of the number of learners by gender in each grade. In the case of the Free State, where a gender breakdown of enrolment was not available in the Snap data, the gender breakdown of the 2001 ASS for Free State learners was first applied to the Free State Snap Survey data, before an age breakdown was calculated.
One further problem identified in the 2001 ASS data related to the number of ELSEN learners. In six of the nine provinces ELSEN enrolment had been disaggregated equally between male and female learners in each age and grade in each school. Only Limpopo, the Northern Cape and Western Cape had different ELSEN enrolment figures for male and female.

**Census data**

Queries were run on the Census 2001 Community Profile Database. Descriptive statistics were calculated from the South Africa by province and municipality data. Information on present school attendance was obtained for weighted persons for present school attendance and geography by age. Data for those whose present school attendance was classified as ‘none’ or as ‘school’ was used. Information on population numbers was obtained for weighted persons for geography and age by population group and gender. Population for the age cohort 7 to 18 years was extracted.

**Limitations**

This study assessed the number of children who are formally registered at school. It did not provide information on attendance at school, which according to Unesco (2002, p.49) is usually lower than NERs.

While the study aimed to provide a rigorous and accurate analysis of access to schooling, this was subject to the quality of the available data. The process of developing the indicators for this report highlighted problems with the available data in terms of coverage and reliability. The ASS data for 2001 undercounted the number of learners and data had to be filled in from the Snap Survey. With regard to the Census data, according to Statistics SA, the adjusted population figures by province, population group, sex and age in the 2001 census data have 95% confidence limits. Also, preliminary independent demographic analyses of the census results suggest that there is an over-estimation of children aged between 10 and 19 years of age (Statistics South Africa, on line, 2005).