A mathematics vocabulary questionnaire for use in the intermediate phase

Marthie van der Walt, Kobus Maree and **Suria Ellis**
marthie.vanderwalt@nwu.ac.za; kobus.maree@up.ac.za; suria.ellis@nwu.ac.za

Teachers and psychologists need an instrument to assess learners' language proficiency in mathematics to enable them to plan and evaluate interventions and to facilitate best practice in mathematics classrooms. We describe the development of a mathematics vocabulary questionnaire to measure learners' language proficiency in mathematics in the intermediate phase. It covers all the steps from designing the preliminary questionnaire to standardising the final instrument. A sample of 1 103 Grades 4 to 7 Afrikaans-, English- and Tswana-speaking learners in North West Province completed the Mathematics Vocabulary questionnaire (Primary) (MV(P)), consisting of 12 items. We analysed the data by calculating discrimination values, performing a factor analysis, determining reliability coefficients, and investigating item bias by language, gender, and grade. We concluded that there was strong evidence of validity and reliability for the MV(P).

Keywords: intermediate phase; mathematics vocabulary; reliability; standardising instrument/questionnaire; validity

Introduction

The context for our research was twofold. Firstly, as is the case in all developing countries, South Africa urgently needs a mathematically and scientifically literate population to guarantee its survival and success in the post-modern world. Since mathematics, physics and chemistry are today regarded as core life-skills in the post-modern, technological world, success in these subjects may well be an indicator of economic success in the future (Teaching, 2004). Learners throughout the country are finding themselves in situations where high demands are made on them to process mountains of information, to master subject content and to apply their knowledge and skills in everyday situations. South African training institutions are being challenged to assess whether they are training learners to master survival skills, to become lifelong learners and to accept responsibility for the learning process. With the increased emphasis on the acquisition of certain outcomes and skills in an outcomes-based education system, teachers have to assess more than only cognitive skills (McNeir, 1993). Since mathematics is so important in education in South Africa today, research on instruments that have a bearing on mathematics is particularly relevant (Department of Education (DoE), 1995). An assessment instrument, that could help teachers determine learners' acquisition and understanding of the limited technical vocabulary of mathematics, would therefore be widely welcomed.

Secondly, apart from the above challenges, the consequences of the former apartheid education system in South Africa are still "catastrophic" (Kahn,

2004:149) and they continue to hamper the career prospects of (especially) black learners. The South African education system has been undergoing extensive restructuring since the advent of democracy in 1994. For example, an outcomes-based education system was introduced in 1998 and amended in 2001. The system holds that all learners have the ability to succeed and it focuses on the acquisition of knowledge, subject skills, language skills, values and attitudes, unlike the traditional practice that was based on content mastery only. In terms of this paradigm, teachers are expected to introduce real-life mathematics into classrooms and help learners acquire the necessary vocabulary and skills that will prepare them to become life-long learners and critical thinkers. It was also envisaged that this new curriculum (colloquially referred to as Curriculum 2005 and based on principles that incorporate learner centredness, formative assessment, integration and critical thinking) would reflect the values and principles of the new democracy in South Africa (Chisholm, Volmink, Ndhlovo, Potenza, Mohamed, Muller, Lubisi, Vinjevold, Ngozi, Malan & Mphahlele, 2000).

However, the introduction of the outcomes-based education system does not presently appear to be yielding satisfactory results; this can be explained as follows:

Researchers agree that the subject matter knowledge (SMK) of the majority of learners in South Africa is parlous. The PIRLS Study (Mullis, Martin, Kennedy & Foy, 2007), the Third International Mathematics and Science Study in 1995, the Third International Mathematics and Science Study — Repeat (TIMSS-R) (1999), and the Trends in Mathematics and Science Study (2003) found that South African learners time and again performed extremely poorly, if not most poorly of the group of more than 40 participating countries (Howie, 1997; 2001). An analysis of the results revealed that learners experienced problems relating to their limited technical vocabulary of mathematics. For example, when they had to understand word problems, articulate the solutions to the problems and write down the solutions. In short, the South African learners struggled most in dealing with problems involving language. In general, the learners in the studies experienced many problems communicating their answers in the language of the test (English), and they revealed their lack of the basic mathematical knowledge required. South African primary school learners' lack of basic mathematics skills and mathematics language proficiency is of particular concern. In the first systematic evaluation of learners' skills in English, mathematics, and science in 2001, Grade 3 learners achieved a national average of 30% in mathematics (Department of Education, 2002), and, in the follow-up study, Grade 6 learners in 2004 (Grade 3 in 2001) averaged 27%. Grade 6 learners in North West Province achieved an average of 24% (North West Education Department, 2006). Hough and Horne Consultants (2007) also found that a fairly sizeable proportion of top achievers across various language groups were not functionally literate/numerate.

Various reasons for this state of affairs have been put forward, including

those of researchers such as Arnott, Kubeka, Rice and Hall (1997), Howie (2001), Maree and Molepo (1999), and Reynolds and Wahlberg (1992). The reasons include the poor socioeconomic background of learners (little incentive to study at home), lack of appropriate learner support materials, general poverty of school environment, general poor quality of teachers and teaching (including poor subject knowledge and poor motivation), language of instruction (often not the same as learners' mother tongue) and an inadequate study orientation. The influence of affective factors on mathematics performance has also been noted by a number of authors (Maree, Claassen & Prinsloo, 1997). Inadequate performance in mathematics is seldom simply a cognitive phenomenon. Unless language-related factors in particular are taken into account, problems in mathematics may well be approached simplistically (Kassiem, 2004).

Teachers and psychologists alike require an instrument to assess learners' language proficiency in order to plan and evaluate interventions and to facilitate best practice in mathematics classrooms. However despite the clear need for a mathematical language proficiency test to assess learners' acquisition and knowledge of the limited technical vocabulary of mathematics, no such test currently exists for the intermediate phase (Maree, Molepo, Owen & Ehlers, 2005).

Objectives of the study
The following research issues were addressed:
1. The standardisation of a mathematics vocabulary questionnaire with acceptable psychometric properties which teachers can use to assess the proficiency of Grade 4 to Grade 7 learners' basic vocabulary in mathematics.
2. The provision of guidelines for other researchers and psychologists on the design and development of similar instruments in developing country contexts.

Research methods
The empirical investigation was divided into two phases: developmental and implementation.

Method in the developmental phase
The action plan for the design of a new multiple-choice measuring instrument, proposed by Nunnally and Bernstein (1994), was considered appropriate for this study. First, we decided on the type of items and the level of difficulty of the items, based on the requirements contained in the learning outcomes for the learning area mathematics in the NCS (DoE, 2003). We also sought the opinions of experts (mathematics teachers, Grades 4 to 7 learners and university teachers), conducted a pilot study, analysed the data statistically (e.g. calculated reliability coefficients, performed item analyses, and calculated item bias and discrimination values), adapted and changed some items, and

discarded/excluded others, to improve the reliability and validity of the original instrument.

Method in the implementation phase
Sampling strategy
A sample of schools selected from the database provided by the North West Education Department and representing the population of North West Province, according to socioeconomic status, language, grade and location (city, town or township), participated in the study. To ensure representation of each significant part of the population, the population was divided into the following strata or subpopulations: gender (male or female), mother tongue (Afrikaans, English, or Tswana), Grade (4 to 7) and area (four distinct regions in North West Province). Schools with fewer than 200 learners were excluded because of their inaccessibility.

Data collection procedures
A sample of 1 103 Grades 4 to 7 learners in North West Province participated voluntarily in the main study. Data were collected by means of the MV(P), which the learners completed. The learners were told that this questionnaire aimed at assessing their mathematics language proficiency in the vocabulary commonly required in Grades 4 to 7 and that they should fill in the multiple-choice questionnaire, individually, within 15 minutes. The learners received a biscuit during the break that preceded the completion of the questionnaire.

Rationale for determining weights to calculate weighted estimators for the population
Because most people in North West Province are Tswana-speakers, we decided not to use a proportional stratified sample in which the ratio of learners per language group was the same as in the population. The much smaller number of Afrikaans and English learners in the sample would have made comparisons on a language basis impossible. We decided instead to use weighted averages for comparing the groups to ensure proportional representation of language. Almost even numbers of learners from each language group were observed so that we would have enough data to do a factor analysis of the data for each language group and to estimate mean values of groups.

Results reported in this study are based on data weighted according to language, so that the final results could be in the ratio in which learners were present in the language groups in the province. The results of the 2001 census are given according to age groups, and the proportion of learners in the different language groups, aged 5 to 14 years, was used to calculate the weights (Table 1).

Some of the initially selected schools were unable to participate in the study due to the prolonged civil servants' strike in 2007. Schools with similar populations to the initially selected schools were invited to participate instead. The Tswana-speaking schools' involvement, especially in Region 4, was affected by the strike.

Table 1   Frequencies of children in the age group 5–14 years, during 2001 census, per language group

	Afrikaans	English	Tswana
Total number	45,442	7,280	554,051

The following schools were involved: three Afrikaans-speaking schools (A1, A2, and A3) from three different regions (1, 2, and 3) in North West Province, one located in a town, another located in a city and another located next to a township/informal settlement; three English-speaking schools (E1, E2, and E3) from Regions 2 and 3 in North West Province, one in a city, one in a town, and one in a township/informal settlement; two Tswana-speaking schools (T1 and T2) from Region 2 in North West Province (both located in townships/ informal settlements).

Between 100 and 187 learners per school participated in the study, totalling 1 103 learners. The percentages of learners from each language group (Afrikaans, English, and Tswana) who participated were 33%, 36%, and 31%, respectively. According to the data retrieved from the 2001 census (Table 1), the calculated weights for the language groups Afrikaans, English and Tswana were 122.8, 18.3, and 1 653.9, respectively. The numbers of male (538) and female (555) learners were almost equal, and the frequency of learners per grade ranged between 267 and 289.

Statistical analysis of data and testing for reliability and validity of the MV(P)
The procedures for analysing the data statistically focused on the standardisation of the MV(P). Reliability and validity were examined to facilitate quality assurance of the developmental phase of the instrument. Statistical techniques included calculation of discrimination values, factor analysis, reliability coefficients, two-way frequency tables, ANOVA, effect sizes (Ellis & Steyn, 2003) as well as Rasch model *p* values to determine item bias, and Tucker's φ coefficients for bias of the instrument between language groups.

Results
Results of the developmental phase
The content of the pen-and-paper measuring instrument was based on the vocabulary for Grades 4 to 7 by the five learning outcomes in the mathematics learning area, as spelled out in the NCS (DoE, 2003). The level of difficulty of items was distributed across the Grades 4 to 7 content of the mathematics learning area as recommended by Scheepers (1992). Multiple-choice items contained five possible responses of which only one was correct. To limit the extent of what learners were expected to read, interpret, and react to, the items were selected according to the developmental levels of Grades 4 to 7 learners. The questions were asked in clear and concise language and the distractors were acceptable and not misleading (TIMSS, 2003). Expert opinion

was sought on the items included in the original instrument: six practising mathematics teachers in Potchefstroom and 10 Grades 4 to 7 learners from all three language groups (Afrikaans, English and Tswana) were requested to indicate their understanding of the items. Various mathematics experts at different universities were also asked to comment on the clarity, simplicity, unambiguity and use of words with exact meanings and the equivalence of the Afrikaans, English and Tswana questions and distractors. Changes were then made to the items where necessary. A total of 353 learners (98 Tswana-, 156 English- and 99 Afrikaans-speaking learners) participated in the pilot study, conducted during November 2006. The learners received something to eat before completing the questionnaire as hunger may have negatively affected the concentration of some learners. The learners were requested to complete the questionnaire and circle statements, words, or parts of sentences they did not understand. The time taken to complete the questionnaires had to be limited (Scheepers, 1992), and the average time taken by the learners in the pilot study to complete the 24 items provided us with a time limit guideline for the completion of the 12 final items in the MV(P). Based on learners' reactions, the formulation of some items was again refined.

Statistical data analysis then followed. We analysed the items according to the learners' responses to each item and to the relationship between the items and the learners' overall performance (Nunnally *et al.*, 1994). Items with discrimination values less than .35 did not discriminate well and were discarded (Nunnally *et al.*, 1994).

Biased items are items in which the scores of learners from the different subgroups (e.g. language/gender/grade) differed (Osterlind & Martois, 1980). Because the pilot study did not yield enough data (the group was too small) to investigate item bias using factor analysis, effect sizes for two-way frequency tables were calculated to determine the practical significance of the relationship between items and the language, gender and grade group. This revealed small (ω < .1) (Ellis & Steyn, 2003) effect sizes, indicating that the relationship between the gender, language and grade groups and individual items was generally of no practical significance.

We also calculated Cronbach α values — a value of .70 for the MV(P) was obtained and was regarded as evidence of reliability for the instrument.

The revised, final version of the MV(P) consisted of 12 multiple-choice items with a 15-minute time limit for completion.

## Results of the implementation phase
Ethical issues
Written permission/consent was requested and received from the North West Education Department authorities in Mafikeng, the four regions in North West, the principals of the participating schools, and the parents of participating learners.

Skewness of items of the MV(P)

The percentage of learners who chose each distractor ranged between 4.5% and 60%, implying that no distractor was chosen by 95% or more of the learners (Frijda & Johada, 1966; Smith, 1988). The results indicated that each distractor was functional because it was chosen by a satisfactory percentage of learners.

Learners were classified into three groups in accordance with their total MV(P) score. The top and bottom 27% of learners were regarded as high and low performers, respectively. The difference in percentage of learners from the low and high performance groups who answered the items correctly is defined as the discrimination value and is shown in Figure 1.
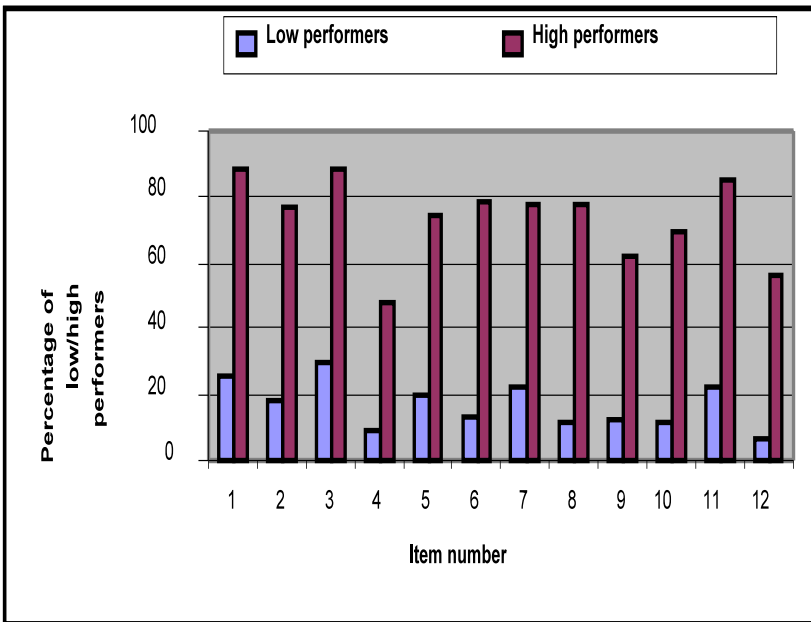
Figure 1 Percentage of respondents in high and low performance groups who answered items in the MV(P) correctly

Figure 1 shows that the learners in the high performance group performed better in all the items than those in the low performance group. Items 4 and 12 were the only items answered correctly by less than 60% of the high performance learners and by less than 10% of the low performers, indicating that these questions were possibly difficult, unfamiliar or confusing to the learners.

Table 2 Percentage of respondents who answered items correctly, missing data, and discrimination value per item

MV(P) item number	Percentage answered correctly	Missing data	Discrimination values
1	57.6	1.8	.64
2	44.6	2.6	.59
3	59.8	2.0	.59
4	22.6	1.9	.40
5	47.4	3.1	.55
6	46.0	3.3	.65
7	47.0	3.4	.56
8	39.4	3.6	.65
9	34.1	3.5	.49
10	35.7	3.1	.57
11	52.7	4.7	.63
12	25.0	12.8	.50

The percentage of learners who answered the items correctly, the percentage of missing data, and the discrimination values for each item are shown in Table 2.

Inspection of Table 2 indicates that the percentage of missing data in Items 11 and 12, the last two items in the MV(P), probably increased because the learners worked too slowly and could not finish the questionnaire in the allocated time, or they could have been tired; Items 4 and 12 were answered correctly by less than 30% of the learners; and Items 8, 9 and 10 were answered correctly by less than 40% of the learners, even though good discrimination values were reported. Discrimination values larger than .35 indicated that all items had acceptable discrimination (Bachman, 2004).

Validity and reliability of the MV(P)
The sample size in the study was in accordance with recommendations in statistics textbooks, namely, a minimum of 300 learners (respondents) is required for factor analysis (Tabachnick & Fidell, 2001; Comrey & Lee, 1992).

An exploratory factor analysis was done on the responses to the 12 items in the MV(P) to determine the underlying factor structure. The adequacy of the sample was measured as 0.849 by the Kaiser-Meyer-Olkin test (KMO) (Field, 2005). According to Field (2005), this indicates that the 12 items correlated sufficiently for factors to emerge. Kaiser's Criterion (Field, 2005) extracts three factors, with eigenvalues larger than 1, from the 12 items in the MV(P). The total variance explained by these three factors was 44.6%. Since correlations between factors were expected, an oblique rotation (*Oblimin*) was used (Field, 2005). The pattern matrix of the rotated factor analysis and the initial communalities are shown in Table 3.

Table 3 indicates the three extracted factors and shows that all item communalities were larger than .3 (Tabachnick *et al.*, 2001). The factor-

correlation matrix indicated correlations of between .24 and .31, which can be regarded as visible to researchers in practice (Steyn, 2002).

Table 3   Pattern matrix of the factor analysis and the initial communalities of items

Item number and content	Factor 1	Factor 2	Factor 3	Communalities
9 (rounding off)	**.710**			.494
12 (litre, capacity)	**.634**			.481
4 (quotient)	**.574**			.367
10 (symmetrical line)	**.399**			.343
2 (decimal number)		**–.744**		.534
6 (improper fraction)		**–.689**		.513
8 (equivalent fractions)		**–.608**		.446
5 (remainder; division)		**–.549**		.382
1 (sum of)			**.708**	.516
3 (half of)			**.638**	.492
7 (even numbers)			**.608**	.420
11 (digital time)			**.464**	.358

Method of extraction: principle component analysis;
rotation: *Oblimin* with Kaiser-normalising

Analysis of reliability was done on the total score of the MV(P) and a reliability coefficient (Cronbach α) of .75 was obtained, indicating internal consistency for the MV(P) (Nunnally *et al.*, 1994).

Table 4   Reliabilities when individual items are left out, correlations with total scores and Cronbach α values

Item excluded	Correlation with total score	α
1	.41	.73
2	.38	.73
3	.36	.73
4	.34	.73
5	.34	.74
6	.42	.72
7	.32	.74
8	.43	.72
9	.33	.74
10	.41	.73
11	.41	.73
12	.42	.73

Table 4 shows that all items' correlations with the total score were larger than .3 (Clark & Watson, 1995), thus also indicating good discrimination values. The reliability (Cronbach α values) when individual items were omitted

ranged between .72 and .74, thus indicating that all items contributed to the construct (Leedy & Ormrod, 2001).

Further statistical techniques used to investigate the reliability of the MV(P) were Pearson's correlations between items and average inter-item correlation, which should range between .15 and .55 (Clark, Watson & Reynolds, 1995). The correlations between items indicates the degree to which the individual items measure the same construct while the average inter-item correlation indicates the degree to which the instrument as a whole measures the same construct. The correlations between items ranged between .10 and .35, and the average inter-item correlation value for the MV(P) was .20, which also confirms the reliability of the MV(P) (Clark, Watson & Reynolds, 1995).

Statistical techniques used to analyse item bias included Rasch models and effect sizes for two-way frequency tables with weighted frequencies, as well as effect sizes for a two-way ANOVA on items, where participants are divided in 10 groups (levels) according to their total MV(P) score and with language as other factor. Before item bias can be investigated, using Rasch models (Linacre & Wright, 1994; McNamara, 1996; Winsteps, 2006), it must be determined whether the data fit the model. When all items fit the model, it confirms that all items measure the same underlying construct (De Bruin, 2004). The mean square infit and outfit values for the MV(P) items ranged between .91 and 1.26, indicating a good fit to the model (McNamara, 1996) (Table 5).

Table 5   Item bias of the items of the Mathematics Vocabulary Questionnaire (Primary) per language from Mantel-Haenszel *p* values (Rasch model)

Item number	Mantel-Haenszek *p* values		
	Afrikaans and Tswana	Afrikaans and English	English and Tswana
1	**.019**	**<.001**	.358
2	**.019**	**<.001**	.554
3	**.009**	.246	.939
4	**.017**	.067	.273
5	**.009**	.439	.648
6	**.017**	**<.001**	**.002**
7	**<.001**	**.005**	.543
8	**<.001**	**.004**	.771
9	**<.001**	.472	.468
10	**<.001**	.119	.742
11	**<.001**	**.001**	.656
12	**<.001**	**<.001**	.568

p < .05; indicates possible item bias

The Rasch model's *p* values showed item bias with regard to language in Items 1, 2, 6, 7, 8, 11, and 12, Table 5. It is known, however, that small *p*

values, indicating significance, can easily be obtained for a large sample (Hair, Anderson, Tatham & Black, 1998) (Table 6).

Table 6 Item bias of the items of the Mathematics Vocabulary Questionnaire (Primary) from effect sizes for (i) two-way tables with weighted frequencies, per language, gender and grade (ii) ANOVA with regard to language and interaction between language and group

	Two-way frequency tables effect sizes ω			ANOVA (Language) effect sizes η²	
Item No.	Grade	Gender	Language	Group bias according to language	Group bias according to the interaction of language and group (level)
1	.21	−.03	**.24**	.033	.019
2	**.36**	.01	.01	.021	.019
3	**.35**	−.03	.09	.001	.009
4	.18	.05	.09	.001	.017
5	**.37**	.01	.09	.001	.009
6	**.46**	.05	.04	.038	.017
7	**.26**	.03	.16	**.056**	<.001
8	**.39**	−.02	.08	.014	<.001
9	.07	−.01	.09	.019	<.001
10	**.25**	−.01	.15	.043	<.001
11	.14	−.13	.19	**.085**	<.001
12	.07	.04	**.27**	**.080**	<.001
	Effect sizes ω ≈ .3; indicate possible item bias			Effect sizes η² ≈ .06; indicate possible item bias	

The two-way frequency table (with effect sizes ω) indicated (i) medium-sized effects (visible differences) with regard to **grade** in Items 2, 3, 5, 6, 7, 8, and 10, (ii) no differences with regard to **gender** in the items of the MV(P), and (iii) medium-sized effects (visible differences) with regard to **language** in Items 1 and 12, Table 6. The ANOVA effect size indicated that items 7, 11 and 12 showed visible bias with regard to language and none for the interaction of language and level, Table 6. The item bias results helped us decide on further exclusions of/changes in items.

Bias of the factor analysis of the MV(P) with regard to Tucker's φ coefficients
The correspondence of the instrument between language groups was evaluated by a factor congruency coefficient, namely, Tucker's φ (Chan, Ho, Leung, Cha & Yung, 1999; Van de Vijver & Leung, 1997). A good correspondence was expressed with Tucker's φ with values > .90 (Table 7), indicating equivalence of the instrument for language groups (Van de Vijver & Poortinga, 1994).

Table 7 Tucker's φ coefficients for the MV(P)

Language	Tucker's φ coefficients
Tswana	.91
English	.99
Afrikaans	.99

Discussion

Discussion of the implementation phase

To the extent that it is possible to draw conclusions about the learners' socio-economic background, in terms of their specific schools according to the **areas** where the learners lived and the availability of **facilities** at the schools, the following conclusions seem plausible: the learners from school A1 were potentially from high socioeconomic backgrounds, the learners from schools A2, E2, E3, and T2 potentially from high to low middle-class backgrounds, and the learners from schools A3, E1, and T1 potentially from very low socioeconomic backgrounds.

Discussion of the results of the data analysis phase

Descriptive statistics of the MV(P)

Questions that seemed to be difficult (answered incorrectly by the learners) were questions on division and quotient (Item 4), and litres and content (Item 12). Difficulties were expected in Items 4 and 12 because a number of learners had enquired about the meaning of "quotient" and "litre" during the completion of the questionnaire. These items were rephrased.

Similarly, the vocabulary in Items 8, 9, and 10 (which included equivalent fractions, rounding off, and a symmetrical line) may have been unfamiliar to learners, the questions too difficult or the questions ambiguous or confusing. After careful deliberation we decided to change Item 10 to eliminate the possibility of confusion.

The large percentage of missing data in Item 11, Table 2, was probably because learners worked too slowly and could not finish in the allocated time, or they could have been tired.

Discussion of the validity and reliability of the MV(P)

The relatively small percentage of variance explained by the three factors (44.6%) could be a result of the fact that the learners completed the questionnaire during June and July and not towards the end of the school year, implying that the items could have been too difficult for some of them. The difficulty in interpreting the three resulting factors (Table 3) theoretically, may have been due to the skewness of items (such as 4, 9, 10 and 12), a result of the level of difficulty of the items and not because of real correlations between these items. This finding as well as the significant correlation between all factors persuaded us to use only a total score for the instrument.

The correlations between items and the average inter-item correlation confirmed that the items measured the same construct, and the value of the Cronbach α coefficient for the MV(P) confirmed the reliability of the MV(P).

Item bias
Visible differences in items (2, 3, 5, 6, 7, 8, and 10) with regard to **grade**, shown in the two-way frequency table, were expected because the Grade 4 learners had been in the intermediate phase for only six months. After careful deliberation, some of the items which showed possible differences with regard to **language** were changed.

Tucker's φ coefficient for the MV(P) indicated very good values (> .9) and confirmed the equivalence of the instrument with regard to language, namely, Afrikaans, English, and Tswana (Van de Vijver & Poortinga, 1994).

Conclusions with regard to the Mathematics Vocabulary Questionnaire (Primary)
We concluded that the MV(P), a 12-item measuring instrument, assessing mathematical language proficiency and based on the requirements in the learning outcomes for the mathematics learning area in the NCS (DoE, 2003), has sound psychometric properties, i.e. content validity as well as construct validity, for the three language groups together (Afrikaans-, English- and Tswana-speaking learners in Grades 4 to 7 in North West Province). Strong evidence for the reliability of the MV(P) was indicated.

Limitations of the study
Purposefully selected samples (instead of random samples) were used. Because we could not prove that the samples were representative of the population, generalisation was impossible. Schools with fewer than 200 learners in North West Province were excluded. Furthermore, on account of the civil servants' strike (June – July 2007), two schools were replaced by alternative schools, which meant that no school from Region 4 in North West Province was included. Tswana-speaking schools were thus selected from the same region. These issues could influence the generalisation value of the study.

Use of the questionnaire and future implications
Although a limited population was used in the study, namely, Afrikaans-, English- and Tswana-speaking learners in Grades 4 to 7 in North West Province, we concluded that the MV(P) can be used — with caution — in other provinces in South Africa. Furthermore, the MV(P) can be included in further research, e.g. pre- and post-interventions, to help teachers assess and detect language proficiency problems.

This instrument, of which the development has been described in this article, may fill the need for an instrument to assess learners' language proficiency which could help teachers and psychologists plan, implement, and assess intervention timeously. While we neither claim to have solved the problem of inadequate achievement in mathematics nor to have developed an instrument that will provide instant solutions to all language-related problems

in mathematics, we nonetheless hope that this study will contribute to the current debate on the teaching and learning of mathematics in developing countries. Further research and implementation of the questionnaire are essential.

Acknowledgements

References

Arnott A, Kubeka Z, Rice M & Hall G 1997. *Mathematics and Science teachers: demand, utilisation, supply, and training in South Africa*. Craighall: Edusource.

Bachman LF 2004. *Workbook for statistical analysis for language assessment*. Cambridge: Cambridge University Press.

Chan W, Ho RM, Leung K, Cha DK-S & Yung Y-F 1999. An alternative method for evaluating congruence coefficients with Procrustes rotation: a bootstrap procedure. *Psychological Methods*, 4:378-402.

Chisholm L, Volmink J, Ndhlovu T, Potenza E, Mohamed H, Muller J, Lubisi C, Vinjevold P, Ngozi L, Malan B & Mphahlele M 2000. A South African curriculum for the twenty-first century: Report of the Review committee on Curriculum 2005. Presented to the Minister of Education, Professor Kader Asmal (Pretoria).

Clark LA, Watson D & Reynolds S 1995. Diagnosis and classification in psychopathology: challenges to the current system and future directions. *Annual Review of Psychology*, 46:121-153.

Comrey AL & Lee HB 1992. *A first course in factor analysis*, 2nd edn. Hillsdale, NJ: Erlbaum.

De Bruin GP 2004. Problems with the factor analysis of items: solutions based on item response theory and item parcelling. *South African Journal of Industrial Psychology*, 30:16-26.

Department of Education (DoE) 1995. The white paper on education and training. *Government Gazette, No.16312*. March. Pretoria: Government Printer.

Department of Education 2002. *Draft education for all status report 2002: South Africa, incorporating country plans for 2002 to 2015*. Pretoria: Government Printer.

Department of Education 2003. *Revised National Curriculum Statement Grades R-9 (Schools). Teacher's guide for development of learning programmes, foundation phase*. Pretoria: Government Printer.

Ellis SM & Steyn HS 2003. Practical significance (effect sizes) versus or in combination with statistical significance (*p* values). *Management Dynamics*, 12:51-53.

Field A 2005. *Discovering statistics using SPSS,* 2nd edn. London: SAGE Publications.

Frijda N & Johada G 1966. On the scope and methods of cross-cultural research. *International Journal of Psychology*, 1:109-127.

Hair JR, Anderson RE, Tatham RL & Black WC 1998. *Multivariate data analysis*. Prentice-Hall, Inc.

Hough & Horne Konsultante 2007. *'n Vergelykende studie: die taalvaardighede (Engels) en syfervaardighede van uitsoekmatrikulante asook veronderstellings en vrae oor regstellende aksie wat daaruit voortvloei*. Available at http://info@tjhorne.co.za. Accessed 10 Oct. 2007.

Howie S 1997. *Mathematics and science performance in the middle school years in South Africa*. Pretoria: Human Sciences Research Council.

Howie S 2001. *Mathematics and Science performance in Grade 8 in South Africa*. Pretoria: Human Sciences Research Council.

Kahn M 2004. For whom the school bell tolls: Disparities in performance in senior certificate mathematics and physical science. *Perspectives in Education*, 22:149-156.

Kassiem A 2004. Grade 6 pupils can't read or write — poll. *Pretoria News*, 26 May.

Leedy PD & Ormrod JE 2001. *Practical research: planning and design*. 7th edn. NJ: Merrill Prentice Hall.

Linacre JM & Wright BD 1994. Reasonable mean square fit values. *Rasch Measurement Transactions*, 8:370.

Maree JG, Claassen NCW & Prinsloo WJ 1997. *Manual for the Study Orientation Questionnaire in Mathematics*. Pretoria: Human Sciences Research Council.

Maree JG & Molepo JM 1999. The role of mathematics in developing rural and tribal communities in South Africa. *South African Journal of Education*, 19:374-381.

Maree JG, Molepo JM, Owen JH & Ehlers R 2005. ’n Probleemgebaseerde benadering tot wiskunde in grade 9 en 11 in die Limpopo-provinsie. *SA Tydskrif vir Natuurwetenskap en Tegnologie*, 24:124-133.

McNamara TF 1996. *Measuring second language performance*. London: Longman.

McNeir G 1993. *Outcomes-based education*. Available at http://eric.uoregon. edu/publications/digests/digest085.html.

Mullis IVS, Martin MO, Kennedy AM & Foy P 2007. *Progress in international reading literacy study. Pirls 2006 report*. Lynch School of Education, Boston College, Chestnut Hill, MA: TIMMS & PIRLS International Study Centre. International Association for the Evaluation of Educational Achievement (IEA).

North West Education Department 2006. *Grade 6 Intermediate phase: Systemic evaluation. Provincial Report*. Pretoria: Government Printer.

Nunnally JC & Bernstein IH 1994. *Psychometric theory,* 3rd edn. New York: McGraw-Hill.

Osterlind SJ & Martois JS 1980. *Latent trait theory applications to test item bias methodology. Research memorandum No. 1*. ERIC — Education Resources Information Centre.

Reynolds AJ & Wahlberg HJ 1992. A structural model of high school mathematics outcomes. *Journal of Educational Research*, 85:150-158.

Scheepers JM 1992. *Toetskonstruksie: teorie en praktyk*. Johannesburg: RAU-Drukpers.

Smith TW 1988. *The ups and downs of cross-cultural survey research. GSS Cross-National Report No. 8*. National Opinion Research Center, University of Chicago.

Steyn HS Jr 2002. Practically significant relationships between two variables. *South African Journal of Industrial Psychology*, 28:10-15.

TIMSS 2003. *TIMSS Assessment frameworks and specifications*. Available at http://timss.bc.edu/timss2003i/frameworks.html. Accessed 1 May 2006.

Tabachnick BG & Fidell LS 2001. *Using multivariate statistics,* 4th edn. Boston: Allyn and Bacon.

Taylor N & Vinjevold P (eds) 1999. *Getting learning right: Report of the President’s Education Initiative Research Project*. Johannesburg: Joint Education Trust.

Teaching 2004. *Beeld*: 9 November. Available at http://www.news24.com/Beeld/Hoofartikels/0,3-65_1618373,00.html. Accessed 9 November 2004.

Van de Vijver FJR & Leung K 1997. Methods and data analysis of comparative research. In: JW Berry, YH Poortinga & J Pandey (eds). *Handbook of cross-cultural psychology, 2nd edn,* 1:257-300. Boston: Allyn and Bacon.

Van de Vijver FJR & Poortinga YH 1994. Methodological issues in cross-cultural studies on parental rearing behavior and psychopathology. In: C Perris, WA Arrindell & M Eisemann (eds). *Parental Rearing and Psychopathology.* Chichester: Wiley.

Winsteps 2006. Winsteps© Rasch Measurement, Version 3.63.2, Copyright© John M Linacre. Available at http://www.winsteps.com

Authors

Marthie van der Walt is Lecturer in school mathematics (foundation, intermediate and senior phases) at the North-West University, Potchefstroom Campus. Her research interests are metacognition in the teaching and learning of mathematics and study orientation in mathematics.

Kobus Maree is Professor in the Faculty of Education, University of Pretoria. A triple doctorate, he is an award-winning and internationally recognised researcher, particularly in career counselling. He is the Editor of *Perspectives in Education* and the *South African Journal of Psychology*.

Suria Ellis is Head Subject Specialist at the Statistical Consultation Service of the North-West University. She is a widely published researcher and the Statistics Editor of *Perspectives in Education.*