COMPARATIVE ANALYSIS OF THE RELIABILITY OF MISSING SCORES IMPUTATION TECHNIQUES

Boipuso Mosalakgotla
Directorate of Research & Policy Development
Botswana Examinations Council
bmosalakgotla@bec.co.bw

Pauline Masemola
Umalusi Council
Pauline.Masemola@umalusi.org.za

Abstract
Examination boards commonly face problems of missing marks during examinations. Examinations data may go missing due to various reasons including hectic logistics and adverse conditions under which the examination was written. In educational assessments, a number of methods have been used to address this problem but their reliability has not been fully explored. This study sought to compare the reliability of five techniques used by members of the Southern Africa Association for Educational Assessment (Regression Analysis, Criterion Mean Method, Same Percentile Position, Z Score Method, Standard Mark Calculation) and another technique used in the United Kingdom (Absolute Standard Deviation Method) in order to recommend a more valid, reliable and fairer technique. The study used Botswana General Certificate of Secondary Education (BGCSE) data from Botswana Examinations Council (BEC) and National School Certificate (NSC) data from Umalusi[1]. Scores were randomly selected, deleted and then estimated using each of the specified techniques. Predicted scores were compared to actual scores using Paired Sample T-test, RMSE and Cohen`s D statistic. The results revealed that Criterion Mean Method (CMM) was superior since precision of its estimated scores was higher. Despite good performance displayed by this method, the study identified limitations which could hinder its full potential. The study developed an improved version of the CMM, tested its performance against the original version and the improved version of CMM is recommended for estimating missing scores.

Keywords: *Examination boards, missing marks, educational assessment, Botswana Examinations Council, Umalusi.*

[1] Council for Quality Assurance in General and further Education and Training

Introduction

Estimation remains paramount in cases of missing data and that does not exclude planning and intervention measures for educational measurements. There are number of planning and interventions carried out through the examinations cycle to ensure that missing data is avoided at all costs. These interventions start with development of an examinations timetable to enable both the examining body and candidates to exercise right and responsibility over the operations of examinations. This minimizes cases of candidates missing examinations papers hence reduces cases of missing data. Despite this there are still cases where candidates miss examination papers due to situations beyond their control or take examination papers in a very bad state. The assessment process is aware that validity of assessment outcomes is associated with these external factors hence the need to deal with missing examinations data to ensure that the results shows a true representation of candidates 'ability. Since in most cases educational assessment comes as a once-off examination and development of an examination follows a complex and time consuming process through which standards of examinations are to be maintained from one paper to another over time, it is cost effective and at the same time fair to estimate the probable true score a candidate would have attained under normal situations. This exercise is done once the examining body has evidence suggesting that the candidate was exposed to unfavorable conditions.

Background to the Study

Many examinations boards around the world are custodians of national examinations results and the results are of high stakes in that they are often used to certify candidates as well as selecting them for higher learning. Due to reasons beyond the control of examination boards, there are often cases of missing scores for some candidates. There are a number of reasons why scores may go missing which includes but not limited to loss of examination papers due to hectic logistics. Some scores may be missing because candidates have taken the examination under adverse conditions such as bereavement and hospitalization. Scores for such candidates may be estimated to give a candidates benefit of doubt; that is if estimation shows that the candidates were disadvantaged by the conditions under which they wrote examinations, their scores will be elevated but if they performed beyond expectations despite the adverse condition under which they wrote examinations, then their scores will be maintained. All efforts are made to ensure that these cases are resolved before the release of examinations results. Examinations boards do estimate scores, but it is apparent that members of Southern Africa Association for Educational Assessment (SAAEA) use different methods of estimating missing scores. It is against such a background that methods of score estimation used by members of SAAEA members were compared to determine which method best estimate missing scores.

Statement of the Problem

Examinations bodies differ on how they deal with missing scores within their databases. Members of SAAEA use different techniques to estimate missing scores and five countries including Botswana and South Africa have been identified. There is a gap in the sense that there is no empirical evidence on the benefit of using one technique over others hence the study would like to compare some of these imputation techniques with the intention of recommending the most valid, reliable and fairer one.

Aims and Objectives of the Study

The study is guided by the following two objectives;

1. To compare the predicted scores obtained by each of the different imputation techniques to the actual scores.
2. To identify the most reliable technique for estimating missing scores.

Literature Review

Rossi et al. (1987) indicated that the best way to solve the problem of missing scores is to avoid missing scores. However much as there are efforts not to we still experience the problem of missing information which may occur for reasons that are beyond control, Piggott et al. (2001). So in such cases, the best way is to treat such scores as missing data and estimate them.

A study similar to this one was carried out in Malawi by the Malawi National Examinations Board (MANEB) where regression model was compared to a number of methods among them Boot Strap method, Z-Score method and mean of criterion variable method. The study revealed that the mean of the criterion variable method of predicting missing scores was better than the other methods used in the study and regression analysis became second. The study recommended that mean of the criterion variable method should be adopted when predicting missing scores from other scores.

Another study was conducted by Smits, N of University of Amsterdam in 2003 which aimed at identifying an appropriate method for estimating missing marks for students at the same university. Theoretic and empirical differences between GPA and 7 alternative missing grade techniques were considered. These 7 techniques were subject mean substitution, corrected subject mean, subject correlation substitution, regression imputation, expectation maximization algorithm imputation, two multiple data imputation methods mainly stochastic regression imputation and data augmentation. The missing grade techniques differed greatly. Data augmentation and stochastic regression imputation appeared to be superior as missing grade technique and these two methods were better than using Grade Point Average (GPA).

In 2009, Marlin, B (2009) of University of Toronto, carried out a study entitled "Missing Data Problems in Machine Learning." which aimed at identifying the best method for predicting non-random missing data in a machinery production set-up. The researcher employed a variety of probabilistic models including finite mixture models, Dirichlet process mixture models, and factor analysis. The results showed that each proposed method achieves a substantial increase in rating prediction performance compared to models that assume missing ratings for missing at random.

In a study entitled "Analysing Data Sets with Missing Data: An Emperical Evaluation of Imputation Methods and Likelihood-Based Methods" by Myrtveit, Stensrud and Olsson (2001) compared four missing data techniques in the context of software cost modelling: Listwise deletion (LD), mean imputation (MI), similar response pattern imputation (SRPI) and full information maximum likelihood (FIML). The results suggested that only FIML is appropriate when the data are not missing completely at random (MCAR). Unlike FIML, prediction models constructed on LD, MI and SRPI data set is too small to enable the construction of meaningful regression-based prediction model.

Methodology

The study adopted a quantitative approach where six methods of computing missing data were compared. The study extracted data from Botswana Examinations Council (BEC) and Umalusi databases for 2015 cohorts at Botswana General Certificate of Secondary Education (BGCSE) and National Secondary Certificate (NSC) respectively.

Sampling

A two-stage sampling technique was used in this study. The first stage of sampling was done at subject level where 4 subjects were selected through purposive sampling to ensure that subjects of different structures are included mainly those that are technical in which outlined answers are used in the marking guide and humanities where subjectivity is applied during marking. To ensure that each group was represented, the follows subjects were used:

- BGCSE Mathematics (Mathematics within the BEC data set) – Technical Group
- NSC Mathematics (Mathematics within the Umalusi data set) – Technical Group
- BGCSE English language (English within the BEC data set) – Humanities Group
- NSC Afrikaans (Afrikaans within the Umalusi data set) – Humanities Group

The second stage of sampling was carried out through simple random sampling of candidate scores where a total random sample of 1200 cases was selected.

Procedure for Comparison

To compare the methods of imputation the following procedure was carried out;

For a subject with 2 papers (paper 1 and paper 2), paper 1 was used to predict paper 2. First a table of random numbers was used to select 30 paper 2 scores to be deleted. These scores were then estimated using the different techniques. Every time a prediction was made using a particular method, the predicted scores were compared to the actual scores using Paired sample T-test. A technique which fails to produce predicted scores which are statistically not different from the actual scores was dropped at this stage.

The effect size of each method of prediction was then calculated using the Cohen`s D formula (practical significance). The smaller the effect size the better the technique and the effect size of <0.3 represented a low effect and the effect size of 0.3 to 0.6 represented moderate effect while the effect size greater than 0.6 was regarded as high. Furthermore, the Root Mean Square Error (RMSE) was used to examine the degree of error for each of the prediction techniques which survived the first stage. The technique with a smaller RMSE was regarded as a better technique. The whole procedure was then repeated after reversing the papers i.e. using paper 2 to predict paper 1. After every comparison, the results were discussed in relation to model significance, correlation coefficient, significance of the mean difference, Cohen`s D and RMSE.

Conceptualization of the Methods considered for this Study

Six methods of estimation were compared mainly Regression Analysis, Criterion Mean Method, Same Percentile Position, Z Score Method, Standard Mark Calculation and Absolute Standard Deviation Method. Scores were deleted and regarded as missing before they were predicted using these methods and the results were discussed in relation to how far they were the from the original scores. The following concepts underline procedure for each method;

Criterion Mean Method (CMM)

Example for subjects with two question papers

When one of the scripts of a candidate is lost, the following procedure will apply:

1. Rank the marks of the paper for which the candidate has a script. The marks are ranked provincially/regionally.
2. Identify the marks of candidates that are to a maximum of 5% above and the same percentage below the mark of the candidate whose script is lost, and use the maximum range possible.
3. Identify the marks of these candidates on the second set of scripts where the candidate has a missing script.
4. The average mark of these candidates will be the mark awarded to for the lost script.

Same Percentile Positioning (SPP)

In case an exam script is missing for a learner, it is of great importance for a candidate whose script is missing to have evidence of other written exam papers or components. The principle is to identify a component or combination of components for the same syllabus for which the candidate does have marks, and to award the learner a mark for the missing component that, as nearly as possible, places them the same percentile of the population as they have achieved on the component(s) being used in the calculation. The mark to be awarded must be a whole mark. Where the relevant percentile occurs between two whole marks, the higher mark should always be awarded.

Z-Score Method (ZSM)

The method is employed for estimating missing script mark as follows:

Where,

$$\mathbf{M_1} = \left[\frac{M_2 - \bar{X}_2}{\sigma_2}\right] \ \mathbf{X}\ \boldsymbol{\sigma_2} + \ \bar{X}_2$$

$\mathbf{M_1}$ = Candidate's mark to be predicted;

$\mathbf{M_2}$ = Candidate's mark in paper that correlates best with paper where a mark is missing;

$\mathbf{X_1}$ = Mean of paper for which a mark is to be predicted;

$\mathbf{X_2}$ = Mean of paper that correlates best with paper for which a mark is to be predicted;

$\boldsymbol{\sigma_1}$ = Standard deviation of paper in which mark is to be predicted.

$\boldsymbol{\sigma_2}$ = Standard deviation of paper that correlates best with paper in which mark is to be predicted.

Regression Analysis (RA)

Regression analysis is employed for estimating a missing mark as follows:

Y = A X + C

Where,

Y = Leaner's mark to be predicted;

X = Learner's mark in paper that correlates best with paper in which mark is to be predicted;

C = the intercept

A= represents the slope or gradient.

Standard Mark Calculation (SMC)

This technique mainly aims at predicting a syllabus grade for a candidate with a missing score in one of the components for the concerned syllabus. The process allows subjectivity on the estimated score to identify a grade suitable for the candidate. The applicability of the technique is illustrated as below;

Calculation of Standard Marks for Rank Order Listing

Syllabus Grade	Standard Mark Threshold	Syllabus Option Grade Threshold (figures shown are for exemplar purposes)
NSSC(O)		
	100	240 (maximum mark for syllabus)
A*	85	175
A	80	166
B	70	147
C	60	128
D	50	109
E	40	90
F	30	71
G	20	52
U	0	0
NSSC(H)		
	100	120
1	85	88
2	80	83
3	70	73
4	60	64
U	0	0

To calculate a standard mark:

Take the candidate syllabus mark and deduct the lower grade threshold boundary mark for the grade awarded.
Then multiply by the difference between the equivalent standard mark threshold boundaries.
Then divide by the difference between the equivalent grade threshold boundaries.
Then add the lower standard mark grade boundary.

Worked example using the exemplar grade threshold boundary marks in the table above:

Candidate syllabus mark was **160** equating to Grade B(b).

160 – 147(lower grade threshold boundary for Grade B) = **13**
× 10(difference between the equivalent boundaries of standard marks i.e. **80** minus **70**) = **130**
÷ 19(difference between grade threshold boundaries i.e. **166** minus **147**) = **6.8**
+ 70(lower standard mark boundary) = **76.8**

76.8 = Standard Mark

<u>Absolute Standard Deviation Method (ASDM)</u>

The Joint Council for Qualifications is responsible for examinations in most of the UK countries including Scotland and Northern Ireland. The method used by JCQ places a candidate at a number of standard deviations below or above the mean where the learner has a mark. Moreover, the performance of leaners where a learner has a missing script is taken into consideration. The standard deviation and the mean are computed where a learner has a missing script. Then from the comparison of standard deviations, the missing script is calculated.

Data Analysis

Because of the limited resources and time constrains, the study took a block analysis approach to maximize available data to derive findings of high reliability and analysis was designed such that at each of the blocks the study will be in a position to drop some imputation techniques looking at their performance in relation to others and the following table provide the structure and purpose of blocks designed for this analysis:

Table 1: Analysis Plan for the Project

Analysis Block 1	Using BGCSE Mathematics & BGCSE English Language	Select 2 Methods
Analysis Block 2	Using NSC Mathematics	Select 1 Method
Analysis Block 3	Using NSC Afrikaans	Final Method Modification

Source: Missing Scores

In Analysis Block 1, 600 cases of candidates` scores were used for analysis and 4 methods of imputation techniques were dropped at this stage to remain with only two techniques which managed to acquire positions 1 and 2 out of a total of 6.

At Analysis Block 2, 300 cases of candidates` scores were used for analysis to compare imputation techniques which acquired positions 1 and 2 at Analysis Block 1. The intention is to recommend a technique which is more appropriate for estimating missing scores.

Analysis Block 3 was suggested mainly to help modify the recommended model to provide the best version of it in estimating missing scores. The study took in to account the possibility of limitations to be identified for each technique and it was appropriate to adjust for the limitations and test the performance of a technique against its improved version before recommending. Hence Analysis Block 3 will compare the best technique with its improved version to check if the improved version is performing better than the original version.

Results and Discussions

In this chapter the study examined the quality of the data and ensured that the data met standards required by all of the imputation techniques to allow estimation required for this comparative analysis.

Comparative Analysis of Techniques

Block Analysis 1

In block analysis 1, 50% of the data (600 cases) were used to identify two methods which consistently outperformed other methods. This was done through the use of data from BEC database in the form of 2015 BGCSE Mathematics results and 2015 BGCSE English results.

Results for BGCSE Mathematics Papers

Results when Mathematics Using Paper2 to Predict Paper1

Table 2: Comparison of Techniques when Mathematics Paper2 Predicts Mathematics Paper1

Statistics	*RA*	*CMM*	*ZSM*	*SPP*	*ASDM*	*SMC*
R	0.90	0.96	0.95	0.95	0.91	0.45
Paired T-test_2 tailed Sig	0.81	0.92	0.77	0.09	0.11	0.33
CI	(-2.20,1.74)	(-1.30,1.44)	(-1.57,1.17)	(-2.72,1.87)	(-3.41,0.34)	(-103.2,36.09)
RMSE	5.19	3.60	3.61	4.03	5.17	186.56
Cohen`s D	0.04	0.02	0.05	0.33	0.30	0.18

Source: Missing Scores; *statistical tests @ α=0.05*

The results shows a paired T-test of 0.81, 0.92, 0.77, 0.09 and 0.33 for RA, CMM, ZSM, SPP, ASDM and SMC respectively hence since all P-values are greater than α=0.05 and we fail to reject H0; the difference between predicted scores and original scores is not statistically significant and conclude that all techniques have successful estimated the missing scores. It can also be observed that despite reliably low Cohen`s D for RA, CMM, and ZSM methods, the CMM outperformed other methods with the smallest RMSE of 3.60 followed by ZSM (3.61) and RA (5.19). SPP and ASDM became 4th and 5th respectively because of the moderate effect size (0.3 to 0.6). SMC became 6th because of very low goodness of fit (r^2 =0.20) which makes its Cohen's D unreliable and also that the

Results for when Using Mathematics Paper1 to Predict Paper2

Table 3: Comparison of Techniques when Mathematics Paper1 Predicts Mathematics Paper2

Statistics	*RA*	*CMM*	*ZSM*	*SPP*	*ASDM*	*SMC*
r	0.96	0.95	0.95	0.95	0.92	NA
Paired T-test_2 tailed Sig	0.53	0.45	0.56	0.70	0.74	NA
CI	(-1.34,2.54)	(-1.47,3.20)	(-1.63,2.97)	(-1.61,3.01)	(-0.25,3.49)	NA
RMSE	5.14	6.21	6.09	6.12	7.89	NA
Cohen`s D	0.12	0.14	0.11	0.11	0.06	NA

Source: Missing Scores; *statistical tests @ α=0.05*

The results shows a paired T-test of 0.53, 0.45, 0.56, 0.70 and 0.704 for RA, CMM, ZSM, SPP and ASDM respectively hence since their P-values are greater than α=0.05 and we fail to reject H0; the difference between predicted scores and original scores is not statistically significant and conclude that these techniques have successful estimated the missing scores. It can also be observed that these methods have low and reliable Cohen`s D statistic, hence RA method outperformed other methods with the smallest RMSE of 5.14 followed by ZSM (6.09), SPP (6.12), CMM (6.21) and ASDM (7.89). <u>SMC failed to estimate Paper2 using Paper1 since the paper not missing constitute to <50% of the syllabus total score and this is a very critical condition for using this method</u>

Results for BGCSE English Language Papers

Results for when Using English Paper2 to Predict Paper1

Table 4: Comparison of Techniques when English Language Paper2 Predicts English Language Paper1

Statistics	*RA*	*CMM*	*ZSM*	*SPP*	*ASDM*	*SMC*
R	0.63	0.49	0.41	0.39	0.42	0.05
Paired T-test_2 tailed Sig	0.25	0.44	0.47	0.83	0.28	0.00
CI	(-0.71,2.57)	(-1.12,2.52)	(-1.39,2.92)	(-1.92,2.38)	(-1.87,2.25)	(30.93,36.92)
RMSE	4.42	4.85	5.74	5.67	6.21	[redacted]
Cohen`s D	**0.21**	**0.14**	**0.13**	**0.04**	**0.23**	

Source: BEC_Umalusi Collaboration; *statistical tests @ α=0.05*

The results show a paired T-test of 0.25, 0.44, 0.47, 0.83 and 0.28 for RA, CMM, ZSM, SPP and ASDM respectively hence since these P-values are greater than α=0.05 and we fail to reject H0; the difference between predicted scores and original scores is not statistically significant and conclude that these techniques have successful estimated the missing scores. It can also be observed that these method have low and reliable Cohen`s D statistic, hence RA method outperformed other methods with the smallest RMSE of 4.42 followed by CMM (4.85), SPP (5.67), ZSM (5.74) and ASDM (6.21). <u>SMC is not comparable to other methods because it failed to derive predicted scores which are statistically not different from actual scores.</u>

Results for when Using English Paper1 to Predict Paper2

Table 5: Comparison of Techniques when English Language Paper1 Predicts English Language Paper2

Statistics	*RA*	*CMM*	*ZSM*	*SPP*	*ASDM*	*SMC*
R	0.53	0.41	0.48	0.43	0.44	0.47
Paired T-test_2 tailed Sig	0.36	0.89	0.41	0.66	0.33	0.000
CI	(-1.09,2.89)	(-1.86,2.12)	(-3.45,1.45)	(-3.18,2.05)	(-2.34,1.99)	(9.48,18.45)
RMSE	5.31	5.24	6.54	6.90	6.98	
Cohen`s D	**0.17**	**0.02**	**0.15**	**0.08**	**0.12**	

Source: BEC_Umalusi Collaboration; *statistical tests @ α=0.05*

The results show a paired T-test of 0.36, 0.89, 0.41 0.66 and 0.33 for RA, CMM, ZSM, SPP and ASDM respectively hence since these P-values are greater than α=0.05 and we fail to reject H0; the difference between predicted scores and original scores is not statistically significant and conclude that these techniques have successful estimated the missing scores. It can also be observed that these method have low and reliable Cohen`s D statistic, hence CMM outperformed other methods with the smallest RMSE of 5.24 followed by RA (5.31) , ZSM (6.54), SPP(6.90) and ASDM (6.98). SMC is not comparable to other methods because it failed to derive predicted scores which are statistically not different from actual scores.

Table 6: Rank of Performance for Methods at Analysis Block Analysis 1

Method	*Positions*	*Performance*	*Overall Position*
RA	3, 1, 1, 1	6	1
CMM	1, 4, 2, 2	9	2
ZSM	2, 2, 4, 3	11	3
SPP	4, 3, 3, 4	14	4
SMC	6, 6, 6, 6	24	6
ASDM	5, 5, 5, 5	20	5

Source: Missing Scores

RA got overall position 1 for this comparison which is an indication that RA is consistently performing above other method for the comparisons done with a positioning combination of (3,1,1,1) followed by CMM, ZSM, SPP, ASDM and SMC with positioning combinations of (1,4,2,2,), (2,2,4,3), (4,3,3,4), (5,5,5,50 and (6,6,6,6) respectively. Therefore RA and CMM methods were compared at block analysis 2 for selection of the final appropriate method for missing scores.

Block Analysis 2

In Block Analysis 2, 25% of the data (300 cases) were used to compare the two methods which consistently outperformed other methods at Block Analysis 1. This was done through the use of data from Umalusi database in the form of 2015 NSC Mathematics results.

Results for NSC Mathematics Papers

Results for NSC Mathematics when Using Paper 2 to predict Paper 1

Table 7: Comparison of Techniques when NSC Mathematics Paper2 Predicts NSC Mathematics Paper1

Statistics	*RA*	CMM
R	0.93	0.94
Paired T-test_2 tailed Sig	0.53	0.76
CI	(-3.13,6.00)	(-3.59,4.86)
RMSE	12.11	11.13
Cohen`s D	**0.12**	**0.06**

Source: BEC_Umalusi Collaboration; *statistical tests @ α=0.05*

The results show a paired T-test of 0.53 and 0.76 for RA and CMM respectively hence since these P-values are greater than α=0.05 and we fail to reject H0; the difference between predicted scores and original scores is not statistically significant and conclude that these techniques have successful estimated the missing scores. It can also be observed that they have low and reliable Cohen`s D statistic, Hence CMM outperformed RA with the smallest RMSE of 11.13 while RA recorded RMSE of 12.11.

Results for NSC Mathematics when Using Paper 1 to predict Paper 2

Table 8: Comparison of Techniques when NSC Mathematics Paper1 Predicts NSC Mathematics Paper2

Statistics	*Botswana*	*Umalusi*
R	0.91	0.95
Paired T-test_2 tailed Sig	1.00	0.96
CI	(-5.23,5.23)	(-4.17,3.97)
RMSE	13.77	10.70
Cohen`s D	**0.01**	**0.01**

Source: BEC_Umalusi Collaboration; *statistical tests @ α=0.05*

The results show a paired T-test of 1.00 and 0.96 for RA and CMM respectively hence since these P-values are greater than α=0.05 and we fail to reject H0; the difference between predicted scores and original scores is not statistically significant and conclude that these techniques have successful estimated the missing scores. It can also be observed that they have low and reliable Cohen`s D statistic, CMM outperformed RA with the smallest RMSE of 10.70 while RA recorded RMSE of 13.77.

Table 9: Rank of Performance for Methods at Block Analysis 2

Method	*Positions*	*Performance*	*Overall Position*
RA	2, 2	4	2
CMM	1, 1	2	1

Source: Missing Scores

CMM got overall position 1 for this comparison which was an indication that CMM was consistently performing above RA method. It is worth noting that despite a consistent performance displayed by CMM over other methods, it has limitation which affects the level of precision it exhibits;

1) The study was of the view that performance is subject to other factors such as school characteristics hence these factors should be considered to increase precision when estimating scores. A candidate should be compared to candidates exposed to same performance related factors as him/her. So CMM is limited since it estimate at provincial/regional level rather than school level and this might lead to unexplained variation which affects estimation negatively.
2) The CMM consider performance of candidates who are 5% around the performance of the concerned candidates in the paper not missing excluding those candidates who got exactly the same mark as the concerned candidate in the paper not missing. That is the method assumed the concerned candidate will not perform similar to candidates who perform exactly the same as him/her in the paper missing without any justification. The study believes that very valuable information is lost and such candidates should be included to increase precision.

The study has therefore used the stated limitations to develop an improved version of CMM. The improved version was then compared to its original version to check if the improved version is better or not.

Block Analysis 3

In Block Analysis 2, 25% of the data (300 cases) were used to compare the modified version of the technique which got position 1 at Block Analysis 2. This was done through the use of data from Umalusi database in the form of 2015 NSC Afrikaans results.

Final Method Modification using NSC Afrikaans Papers

Results for NSC Afrikaans when Using Paper 2 to predict Paper 1

Table 10: Comparison of Techniques when NSC Afrikaans Paper2 Predicts NSC Afrikaans Paper1

Statistics	***CMM***	***Improved-CMM***
R	0.87	0.87
Paired T-test_2 tailed Sig	0.58	0.43
CI	(-2.04,3.64)	(-3.47,2.27)
RMSE	7.60	7.57
Cohen`s D	**0.11**	**0.08**

Source: BEC_Umalusi Collaboration; *statistical tests @ α=0.05*

The results show a paired T-test of 0.58 and 0.43 for CMM and Improved-CMM respectively hence since these P-values are greater than α=0.05 and we fail to reject H0; the difference between predicted scores and original scores is not statistically significant and conclude that these techniques have successful estimated the missing scores. It can also be observed that they have low and reliable Cohen`s D statistic, Improved-CMM outperformed CMM with the smallest RMSE of 7.57 while CMM recorded RMSE of 7.60.

Results for NSC Afrikaans when Using Paper 1 to predict Paper 2

Table 11: Comparison of Techniques when NSC Afrikaans Paper1 Predicts NSC Afrikaans Paper2

Statistics	*CMM*	*Improved-CMM*
R	0.87	0.89
Paired T-test_2 tailed Sig	0.57	0.57
CI	(-1.81,3.14)	(-2.87,1.61)
RMSE	6.55	5.94
Cohen`s D	**0.10**	**0.11**

Source: BEC_Umalusi Collaboration; *statistical tests @ α=0.05*

The results show a paired T-test of 0.57 for both CMM and Umalusi Improved-CMM respectively hence since these P-values are greater than α=0.05 and we fail to reject H0; the difference between predicted scores and original scores is not statistically significant and conclude that these techniques have successful estimated the missing scores. It can also be observed that they have low and reliable Cohen`s D statistic, Improved-CMM outperformed CMM with the smallest RMSE of 5.94 while CMM recorded RMSE of 6.55.

Table 12: Rank of Performance for Methods at Block Analysis 3

Method	*Positions*	*Performance*	*Overall Position*
CMM	2, 2	4	2
Improved-CMM	1, 1	2	1

Source: Missing Scores

Improved-CMM got overall position 1 for this comparison which is an indication that Improved-CMM was consistently performing above CMM hence Improved-CMM was finally recommended over its original version.

Summary of Findings

At Block Analysis 1, the study dropped SMC, ASDM, SPP and ZSM in this order respectively and the following limitations were identified for each technique:

1. **SMC**: This technique mainly aims at predicting a syllabus grade for a candidate with a missing score in one of the components for the concerned syllabus. Its process allows subjectivity and this escalates error such that in most cases it failed to predict missing scores which were statistically not different from the actual scores.
2. **ASDM**: This technique places a candidate at n standard deviations above or below the mean and does not consider a fraction of a standard deviation hence n is rounded to the nearest whole number. This rounding contributed to some error and the technique produced estimated scores with more error.
3. **SPP**: This technique places a candidate in the same distribution position as that of the paper with a score that is not missing. It assumes that a candidate will perform the same way with respect to other candidates in both papers. Where this assumption does not hold, the technique estimate missing scores with high levels of error.

4. **ZSM**: This technique assumes a standard normal distribution of scores for both independent variable paper and dependent variable paper. It produces more accurate scores when the standard normality assumption holds and less precise scores when the assumption does not hold.

At Block Analysis 2, the study dropped RA against CMM and the following limitations were identified:

1) **For RA**, it is apparent that statistical limitations of a regression function had significant impact against this method. Regression analysis normally turns to underestimate scores for high performing candidates and overestimate scores of low performing candidates since the techniques averages along the regression line.

2) **For CMM**, the study is of the view that performance is subject to other factors such as school characteristics hence these factors should be considered to increase precision when estimating scores. A candidate should be compared to candidates exposed to same performance related factors as him/her. So the CMM is limited since it estimate at provincial/regional level rather than school level and this might lead to unexplained variation which affects estimation negatively. The CMM consider performance of candidates who are 5% around the performance of the concerned candidates in the paper not missing excluding those candidates who got exactly the same mark as the concerned paper in the paper not missing. That is the method assumed the concerned candidate will not perform similar to candidates who perform exactly the same as him/her in the paper not missing without any justification. The study believes that very valuable information is lost and such candidates should be included to increase precision.

   At Block Analysis 3, the study dropped CMM against Improved-CMM mainly because the Improved-CMM consistently produced estimated scores with less error than CMM. The study identified Improved-CMM as a more valid, reliable and fairer technique to be used when dealing with missing scores.

Conclusion and Recommendations

The study concludes that reliability of missing score imputation techniques used by members of SAAEA differs. The most reliable technique is Improved Criterion Mean Method which was developed by taking care of limitations exhibited by Criterion Mean Method. The Criterion Mean Method became second best followed by Regression Analysis, Z-Score Method, Same Percentile Positioning, Absolute Standard Deviation Method and Standard Mark Calculation respectively. The study recommends Improved Criterion Mean Method to be used when estimating missing scores.

References

Bluman, A.G. (1998), *Elementary statistics, a step by step Approach. 3rd Edition*, Community College of Allegheny country. Mc Graw Hill New York.

Gupta, V. (1999). SPSS for Beginners, vgupta1000@aol.com; SPSS.org

Marlin, B. (2008). Missing Data Problems in Machine Learning. Doctor of Philosophy Graduate, Department of Computer Science, University of Toronto 2008.

Marlin, M.M. & Zemel, S.R. (2009). Collaborative Prediction and Ranking with Non-random Missing Data. ACM, 5-12.

Myrtveit, N., Stensrud, E. & Olsson, U. H. (2011). Analysing Data Sets with Missing Data: An Emperical Evaluation of Imputation Methods and Likelihood-Based Methods. IEEE Transections on Software Engineering, Vol. 27, No. 11.

Piggott, T.D. (2001). Review of methods for missing data. Educational Research and Evaluation, 17(4), 353-383.

Rossi, H., Smits, N. and Vorst, H. (1987), Alternative missing data techniques to grade point average: Imputing unavailable grades. Journal of Education Measurement. 39(3). 187-206.

Shanghnessy, J.J. (2003), Research methods in Psychology, 6th Edition, Hope College, Mc Graw Hill, New York.

Willmott, C. J., & Matsuura, K. (2005). Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in Assessing Average Model Performance. Climate research, 30(1):79.