

**THE APPLICATION OF DISCRIMINANT ANALYSIS AND
LOGISTICAL REGRESSION AS METHODS OF
COMPILATION IN THE PREDICTION FUNCTION IN
YOUTH RUGBY**

by

CONRAD BOOYSEN

Presented to fulfill the requirements for the degree

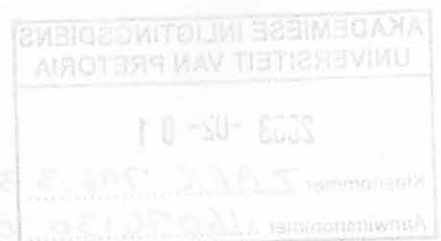
MAGISTER ARTIUM (HMS)

in the

**FACULTY OF ARTS
(Department Biokinetics, Sport and Leisure Sciences)**

University of Pretoria

May 2002



GRATITUDE

I express my gratitude to my Lord and Saviour who gave me the strength and courage to complete this study. I would also like to express my gratitude to the following people for their various contributions:

- My wife, Chantél for her patience and unwavering support and love.
- My parents for making numerous sacrifices for us in their life and without whom being able to study would have been impossible.
- My brother, Anton for his amusing yet effective motivational methods.
- Professor P E Kruger, of the Department of Biokinetics, Sport and Leisure Sciences – University of Pretoria, who acted as promotor responsible for coordinating the requisite academic administration and registration procedures. Thank you for his patience and finest attention to detail.
- Professor E J Spamer of the Faculty of Health Sciences at the Potchefstroomse Universiteit vir Christelike Hoër Onderwys who acted as co-promotor and who provided valuable insight and advice with this dissertation. Thank you for his patience.
- All my friends who provided moral support, encouragement and statistical advice. Thank you for your constant enquiries as to my progress.

SUMMARY

TITLE: The application of discriminant analysis and logistical regression as methods of compilation in the prediction function in youth rugby.

CANDIDATE: CONRAD BOOYSEN

SUPERVISOR: Prof. P E Krüger

DEPARTMENT: BIOKINETICS, SPORT AND LEISURE SCIENCES

DEGREE: M.A. (HMS)

Talent identification is a process where potential sportsmen/women are identified and developed in a specific type of sport. One of the primary reasons for talent identification is to ensure the future survival and development of the particular sport in which the talent is identified and sought.

Another very important reason for talent identification is to be of assistance to the individuals who are being tested for talent. Should these individuals possess talent, they can be further developed. Should they not possess talent for a particular sports-type, they can be redirected in responsible and sensitive manner

Worldwide there is a move toward the identification of talent, with certain countries ahead in the process, successfully applying it to their national sporting codes. Due to our exclusion from international competition for many decades, we find ourselves in the position of being behind in the field of talent identification, albeit catching up very fast.

What has been noticed however, is the move towards the favoring of two primary methods of talent identification, namely discriminant analysis and logistical regression. Different countries of the world use either of these methods, with an open debate currently being fought as to the predictive abilities of these respective methods.

This study compares the two main models of talent identification i.e.: discriminant analysis and logistical regression in terms of their ability to predict talent and their prediction functions, hopefully assisting in resolving the debate as to which is the better model and why.

These models were applied to two U/12 groups of rugby players. One group was talented, the other less so. The results were compared and then certain conclusions and recommendations were made about these models.

A literature study was also done to give this study a philosophical foundation. This included evaluations of other models, the development of these models in history, as well as discussions on whether talent identification is needed or not.

Honest discussions about the advantages, disadvantages and options of combining the two models as well as recommendations for future applications are also contained in this study. The hypothesis is made that there is a significant difference in the predictive function of the one model with relation to the other. This hypothesis has proved to be inaccurate, with the predictive function found to be exactly the same in both models.

Other questions were asked and answered by this study. With the hypothesis being disproved, some of the questions were found to be irrelevant and there was no need to evaluate them further. One question that is investigated and evaluated is whether it is possible to combine these two models to form a more accurate model.

Through various methods of statistical evaluation and substitutions, it is found that by combining the two models a less accurate predictor is formulated, negating the need to combine the two models. This then elicited the conclusion that each model, whilst both being 100% accurate, could be applied under different circumstances when different information is sought.

Discriminant analysis provides an accurate view of the best discriminating factors involved in talent, while logistical regression provides a view of the relative impact of the various factors that determine talent. As can be seen, both of these models can and should be used effectively in the identification of talent.

The proposition is made that when talent is identified, both models be used rather than following an "either-or" approach. The advantage with this is that a well-rounded view of the individuals under scrutiny after applying both models is formulated. Therefore, in summary it can be said that both models predict with the same accuracy, with each model having it's own unique areas of application.

Keywords: Talent Prediction function
 Identification Logistical regression
 National sporting codes Discriminant analysis
 Rugby Discriminating factors
 Methods of compilation Relative impact

OPSOMMING

TITLE:	Die toepassing van diskriminante analise en logistiese regressie as metodes van samestelling in die voorspellingsfunksie in jeug rugby.
KANDIDAAT:	CONRAD BOOYSEN
PROMOTOR:	Prof. P E Krüger
DEPARTEMENT:	BIOKINETIKA, SPORT – VRYETYDWETENSKAPPE
GRAAD:	M.A. (MBK)

Talentidentifisering is 'n proses waardeur potensiële sportlui geïdentifiseer en ontwikkel word in 'n spesifieke sportsoort. Een van die primêre redes vir talent identifikasie is om die toekomstige oorlewing en ontwikkeling van die sportsoort waarin talent geïdentifiseer en gesoek word, te verseker.

'n Ander baie belangrike rede vir talentidentifikasie is om van hulp te wees vir die individu wat getoets word vir talent. Sou die individu oor die nodige talent beskik, kan hulle verder ontwikkel word. Sou die individu nie oor die nodige talent beskik vir 'n sekere sportsoort nie, kan hulle in 'n verantwoordelike en sensitiewe manier begelei word in 'n rigting waarin hulle heelwaarskynlik sou presteer.

Wêreldwyd is daar 'n neiging na die identifikasie van talent, met sekere lande wat ver voor is in die proses, waar hulle dit suksesvol op hulle nasionale sportkodes toepas. As gevolg van die uitsluiting uit internasionale deelname vir etlike dekades, bevind Suid Afrika hom in die posisie waar hy agter is op die gebied van talentidentifikasie, alhoewel daar baie vinnig vordering gemaak word.

Wat baie duidelik is, is die neiging na die voorkeur van twee primêre metodes van talentidentifikasie, naamlik diskriminante analise en logistiese regressie. Verskillende lande in die wêreld gebruik of die een metode of die ander, waar daar huidiglik 'n ope debat is oor die voorspellings vermoë van die twee metodes.

Hierdie studie vergelyk die twee hoof modelle vir talentidentifikasie, naamlik diskriminante analise en logistiese regressie in terme van hulle vermoë om talent te identifiseer en hulle voorspellings funksies

Hierdie modelle was toegepas op twee O/12 rugby groepe. Een groep was talentvol, en die ander een minder talentvol. Die resultate was vergelyk en dan was sekere gevolgtrekkings en aanbevelings gemaak oor die twee modelle.

Die literatuur was ook bestudeer om hierdie studie 'n filosofiese grondslag te gee. Hierdie het evaluering van ander modelle, die ontwikkeling van die modelle in die geskiedenis en 'n bespreking oor of talentidentifikasie nodig is al dan nie, ingesluit.

Eerlike besprekings oor die voordele, nadele en die opsie van 'n kombinasie van die twee modelle sowel as aanbevelings vir toekomstige toepassings is ook ingesluit. Die hipotese is geformuleer dat daar 'n beduidende verskil is in voorspellingsfunksie van die een model t.o.v. die ander model. Die hipotese is bewys om onakkuraat te wees, met die voorspellingsfunksie wat dieselfde is vir albei modelle.

Verskeie vrae is gevra en beantwoord in die studie. Met die hipotese wat bewys is om onakkuraat of onwaar te wees, is van die vrae bevind om irrelevant te wees en is dit nie verder geëvalueer nie. Een vraag wat ondersoek en geëvalueer is, is of dit moontlik is om die twee modelle te kombineer om 'n meer akkurate model saam te stel.

Deur verskeie metodes van statistiese evaluasie en vervangings, is dit gevind dat deur die twee modelle te kombineer, 'n meer onakkurate voorspeller geformuleer is. Die konklusie is gemaak dat elke model 100% akkuraat is, en dat die modelle onder verskillende omstandighede toegepas kan word wanneer verskillende soorte inligting gesoek is.

Diskriminante analise voorsien 'n akkurate oorsig oor die faktore wat die beste diskrimineer (diskriminerende faktore) t.o.v. talent, terwyl logistiese regressie 'n oorsig voorsien van die relatiewe impak van die verskillende faktore wat talent bepaal. Soos gesien kan word, kan en moet altwee die modelle effektief gebruik word om talent te identifiseer.

Die voorstel is gemaak dat wanneer talent geïdentifiseer word, albei modelle gebruik word, in plaas van 'n "of die een of die ander" benadering. Die voordeel hiervan is dat jy 'n volledige heelbeeld van die individu wat onder fokus is, kry. In opsomming kan dit gesê word dat albei modelle met dieselfde akkuraatheid talent voorspel, met elke model wat oor sy eie toepassingsgebied beskik.

Sleutel Terme:	Talent	Voorspellingsfunksie
	Identifikasie	Logistiese regressie
	Nasional sportkodes	Diskriminante analise
	Rugby	Diskriminerende faktore
	Metodes van samestelling	Relatiewe impak

TABLE OF CONTENTS

Title Page	(i)
Gratitude	(ii)
Summary	(iii)
Opsomming	(v)
Table of Contents	(vii)
List of Figures	(xi)
List of Tables	(xii)
CHAPTER ONE: RESEARCH PROBLEM AND GOAL OF STUDY	
1.1 Introduction	1
1.2 Statement of problem	3
1.3 Goal of study	7
1.4 Hypothesis	7
1.5 Methods	7
1.5.1 Literature study	7
1.5.2 Empirical investigation	8
1.5.2.1 Design	8
1.5.2.2 Measuring instruments	9
1.5.2.3 Data analysis/assimilation	11
CHAPTER TWO: REVIEW OF RELATED LITERATURE AND STATISTICAL METHODS	
2.1 Introduction	12
2.2 Background	15
2.2.1 Talent identification-inborn talent and training	19
2.2.2 Models in talent identification	25
• The model of Harre (1982)	27
• The model of Havlicek et al. (1982)	28

• Gimbel's model (1976)	28
• The model of Bompa (1985)	29
2.2.2.1 Conceptual model for talent identification	31
• The Régnier model (1987)	31
2.2.2.1.A Identification of sport specific requirements	32
2.2.2.1.B Identification of determinants of performance	33
• Phases of talent identification	36
Phase 1	36
Phase 2	36
Phase 3	37
2.2.3 Talent identification in South Africa	38
2.2.3.1 Conclusion	41
2.3 Statistical methods	43
2.3.1 Types of multivariable techniques	44
2.3.2 Logistical regression	45
2.3.3 Discriminant analysis	49
CHAPTER THREE: EMPIRICAL INVESTIGATION	
3.1 Phase 1	51
3.1.1 Talent identification in rugby	51
3.1.1.1 Talent identification in youth rugby-a game analysis	51
3.1.1.2 Methods of talent identification in youth rugby players	53
A. Selector selection	53
B. The use of prediction functions	53

3.2 Phase 2	55
3.2.1 A brief discussion of the full test battery	55
3.2.2 In-depth description of the full test battery	56
A. Rugby-specific components	56
B. Motor-physical abilities	58
C. Anthropometric variables	61
1. Kinanthropometric terminology	61
2. Variables, measuring techniques and apparatus	62
CHAPTER FOUR: RESULTS AND DISCUSSION	
4.1 Introduction	66
4.2 Comparison of descriptive statistics of the 2 groups	67
4.2.1 Characteristics of table 3	70
4.2.1.1 Sport-specific skills	70
4.2.1.2 Physical-motor abilities	71
4.2.1.3 Anthropometrical data	71
4.3 Discriminant analysis	73
4.3.1 Comparison of predictive function of the 2 groups	74
4.3.2 Cross-validation of classification function of discriminant analysis	75
4.4 Logistical regression	76
CHAPTER FIVE: CONCLUSIONS AND RECOMMENDATIONS	
5.1 Summary	81
5.2 Conclusions	84

LIST OF FIGURES

5.3 Recommendations

88

Figure 1: Running & catching

88

BIBLIOGRAPHY

91

Figure 2: Agility run

88

LIST OF FIGURES

Figure 1:	Running & catching were used in the study	58
Figure 2:	Agility run using the symbols used in the notation of the language of this study	60
Table 3:	Descriptive statistics and significant differences of the SARFU and MSP groups with regards to rugby-specific skills, motor physical abilities and anthropometrical data	63
Table 4:	Identification of best discriminating factors, using statistical discriminant analysis	67
Table 5:	Coefficients of classification functions of the SARFU and MSP groups	73
Table 6:	Cross-validation of the discriminant analysis	74
Table 7:	Cross-validation of the probability function of logistical regression	78
Table 8:	Explanatory factors found to be best determinants of success in both predictive function methods	79
Table 9:	Table illustrating results obtained when regional regression factors are substituted into discriminant analysis method	80

LIST OF TABLES

Table 1:	Battery of tests that were used in the study	10
Table 2:	Key explaining the symbols used in the notation of the findings of this study	67
Table 3:	Descriptive statistics and significant differences of the SARFU and MSP groups with regards to rugby-specific skills, motor-physical abilities and anthropometrical data	69
Table 4:	Indications of best discriminating factors, using stepwise discriminant analysis	73
Table 5:	Coefficients of classification functions of the SARFU and MSP groups	75
Table 6:	Cross-validation of the categorized discriminant analysis	76
Table 7:	Cross-validation of the probability function of logistical regression	78
Table 8:	Table listing factors found to be best determinants of success in both predictive function methods	79
Table 9:	Table illustrating results obtained when logistical regression factors are substituted into discriminant analysis method	80