

Thesis submitted for the Masters of Education (Honours)

University of Canberra

**An Examination and Critique of the History and Method of
Tertiary Entrance Ranking in the Australian Capital
Territory**

Josephine Dixon

September 2010

Abstract

Since the inception of a separate tertiary ranking scheme for the Australian Capital Territory (ACT) in 1976, several government reports have acknowledged the difficulties of comparing student performance between senior secondary schools in a system that does not have a common, mandated senior secondary school curriculum. For over three decades, these difficulties appear to have been tolerated in deference to a prevailing sentiment of granting schools autonomy in developing and delivering curriculum and assessment at the senior secondary level (years 11 and 12 of schooling). Over that period, there have been several studies and reports investigating bias as a result of the Australian Scholastic Aptitude Test (ASAT),¹ being used as the primary scaling instrument for course² scores both between courses and between colleges.³ The reports have resulted in several changes. Principally, the original ASAT has been expanded to measure a broader base of skills and the linear transformation to ASAT has been replaced by the Other Course Score (OCS) method of scaling. It is generally believed that the OCS scaling method addresses previous concerns by introducing a third set of scores (the v scores) against which students' course scores are thought to be scaled. However, the role of the v scores is relatively minor and course scores continue to be largely scaled to the AST scores. The OCS method was never designed to address, nor has it addressed, many of the biases that have been of concern.

This study explores the evolution of the ACT tertiary ranking system and describes how the current OCS scaling system for ranking students for tertiary entrance operates. Through simulation it investigates the role of the correlation between AST and course/scaling group scores in providing a fair basis for comparing student performance between colleges. Results for cohorts of students are randomly generated with statistical characteristics based on the annual report figures from the ACT Board of Senior Secondary Studies (BSSS). The aggregates are calculated

¹ Now called the ACT Scaling Test – AST.

² Throughout this study, courses and scaling groups are used interchangeably in the context of the ACT system. Scaling groups may have students from more than one course placed on a common scale by the college when the student numbers on one or more of the courses are too low to be scaled alone. In the model, each course is its own scaling group.

³ 'Colleges' refer to year 11 and 12 as they are separate educational institutions in the ACT government system. Although non-government schools do not have their senior years in separate institutions, they are included in the term 'colleges'.

and then recalculated after removing high AST students from the cohorts and then low AST students. After the removal of students, the difference in the students' aggregate is calculated. This process is repeated for three types of cohorts – those with authentic correlations between AST and course scores, those with higher correlations between AST and course score and those with lower correlations. The results indicate that the cohort with whom a student is scaled has a significant impact on the student's final Australian Tertiary Admission Rank (ATAR) and that the impact is lessened with improvements to the AST/course score correlations.

As many senior secondary schools in the ACT now teach a common – albeit not mandated – curriculum in Years 11 and 12 and a national Australian curriculum is likely to be introduced in the near future, this study suggests it would be timely for the ACT government to review its method of ranking students for tertiary entrance.

Acknowledgements

I would like to acknowledge my supervisors, Dr Louise Watson and Dr Judith Ascione, for all of their advice, support and encouragement.

I would also like to acknowledge, Sheridan Roberts, who edited the thesis according to standards D and E of the Australian Standards for Editing Practice (covering clarity of expression, grammar, spelling, punctuation, completeness, formatting, styles, table of contents, and lists of tables and figures).

TABLE OF CONTENTS

Abstract	ii
Certificate of Authorship of Thesis.....	iv
Acknowledgements.....	v
Introduction	xi
Chapter 1. Genesis and History of the ACT Senior Secondary System and Tertiary Ranking Methodology	1
Introduction	1
Before ACT Autonomy.....	1
The Emergence of an ACT Senior Secondary Assessment System.....	3
The First Decade of the ACT Senior Secondary System	12
The Current Era.....	18
Conclusion.....	20
Chapter 2. Review of Assessment Systems of Other Jurisdictions	21
Review of Assessment Systems of Other Jurisdictions	21
New South Wales	21
Victoria.....	22
Western Australia.....	24
Queensland	27
Tasmania	30
South Australia (and Northern Territory).....	31
Chapter 3. The Current ACT System	33
Overview	33
The Initial Scaling Process.....	35
The Other Course Scores (OCS) Scaling Process.....	36
The Iterative Process	42

Chapter 4. Methodology	44
Introduction	44
Construct a Correlation Matrix of Student Course and AST Scores.....	45
Allocate Students to Courses.....	48
Scaling.....	51
Removing Students	55
Chapter 5. Results	57
Generation of the Authentic Cohorts	58
Removing Higher AST Students from the Authentic Cohorts.....	64
Removing Lower AST Students from the Authentic Cohorts	67
Generation of the Higher AST/course Score Correlations Cohorts	70
Removing Higher AST Students from the Higher AST/course Score Correlation Cohorts.....	73
Removing Lower AST Students from the Higher AST/course Score Correlation Cohorts.....	77
Generation of the Lower AST/course Score Correlations Cohorts.....	80
Removing Higher AST Students from the Lower AST/course Score Correlation Cohorts.....	83
Removing Lower AST Students from the Lower AST/course Score Correlation Cohorts	85
Summary of Main Findings.....	87
Chapter 6. Conclusion.....	93
Appendix	96
Bibliography	97

LIST OF TABLES

Table 1. Correlations of TEEP, HSC and teacher predictions with first year ANU results	10
Table 2. Correlations of TEEP tests and HSC subjects	11
Table 3. Comparison of correlations of 1988 scaling methods	17
Table 4. Standardised marks for external examinations	24
Table 5. 2008 Correlations between internal and external assessments for WA and the ACT	27
Table 6. Subject contributions to fields	29
Table 7. Correlations between courses and courses with AST	58
Table 8. Correlations between courses and courses with AST: each of three runs as generated by 'R'	59
Table 9. Correlations between courses and courses with AST: each of three runs after course allocation	61
Table 10. Course score correlations with AST as specified to 'R' as 'R' generated and after students are allocated to classes	62
Table 11. AST means and standard deviations for each course in each run after class allocation	62
Table 12. Components that create the scale and the scaled mean and standard deviation for each course in each run	63
Table 13. Number and percentage of high AST students removed from each course in each run	64
Table 14. Components that create the scale and the scaled mean and standard deviation for each course in each run after high AST students are removed	65
Table 15. Number and percentage of low AST students removed from each course in each run	67
Table 16. Components that create the scale and the scaled mean and standard deviation for each course in each run after low AST students are removed	69
Table 17. AST correlations with each course as specified to 'R' for the high correlation cohorts	70
Table 18. Correlations between courses and courses with AST Each of three runs after course allocation	72
Table 19. Components that create the scale and the scaled mean and standard deviation for each course in each run for high correlations and full cohorts	73
Table 20. Number and percentage of high AST students removed from each course in each of the high correlation runs	74
Table 21. Components that create the scale and the scaled mean and standard deviation for each course in each run for high correlations when high AST students are removed	75
Table 22. Number and percentage of low AST students removed from each course in each run of the high correlation runs	77

Table 23. Components that create the scale and the scaled mean and standard deviation for each course in each run for high correlations when low AST students are removed	78
Table 24. AST correlations with each course as specified to 'R' for the high correlation cohorts	80
Table 25. Correlations between courses and courses with AST each of three runs after course allocation	81
Table 26. Components that create the scale and the scaled mean and standard deviation for each course in each run for low correlations with full cohort	82
Table 27. Components that create the scale and the scaled mean and standard deviation for each course in each run for low correlations when high AST students are removed	83
Table 28. Components that create the scale and the scaled mean and standard deviation for each course in each run for low correlations when low AST students are removed	85
Table 29. Average AST/course score correlations for authentic, higher and lower AST/course score correlation runs	88
Table 30. Summary of ATAR differences	95

LIST OF FIGURES

Figure 1. Mean AST score by year for two colleges	19
Figure 2. Standardised marks for external examinations	25
Figure 3. Object definition diagram for cohorts	50
Figure 4. Object definition diagram for ScaleResults	52
Figure 5. Distribution of AST scores with and without high scores removed	65
Figure 6. Combination of all three runs – some students with high ASTs removed	66
Figure 7. AST scores with and without low scores removed	68
Figure 8. Combination of all three runs – students with low ASTs removed	70
Figure 9. Combination of all three higher correlation runs – students with high ASTs removed	76
Figure 10. Higher correlation and authentic correlation together with high ASTs removed	76
Figure 11. Combination of all three higher correlation runs – students with low ASTs removed	79
Figure 12. Higher correlation and authentic correlation together with low ASTs removed	79
Figure 13. Combination of all three lower correlation runs – students with high ASTs removed	84
Figure 14. Lower correlation and authentic correlation together with high ASTs removed	84
Figure 15. Combination of all three lower correlation runs – students with low ASTs removed	86
Figure 16. Lower correlation and authentic correlation together with low ASTs removed	86
Figure 17. Higher correlation and authentic correlation together with high ASTs removed	89
Figure 18. Higher correlation and authentic correlation together with low ASTs removed	90
Figure 19. Lower correlation and authentic correlation together with high ASTs removed	91
Figure 20. Lower correlation and authentic correlation together with low ASTs removed	91

Introduction

The purpose of this thesis is to investigate the validity of the Other Course Scaling (OCS) that is currently used in the Australian Capital Territory for tertiary entrance ranking. The impact of the students' cohort on their final scores is a particular focus of the study.

In the 1960s, for the first time some programmes at some universities began to have a greater number of applicants than available places. This development necessitated the introduction of quotas and, thus, a method of ranking in order to select successful candidates. During this time all of the Australian senior secondary systems had been gradually moving away from assessment based on 100% final examination to systems of continuous school-based assessments which contributed to at least part of students' matriculation results. In order to provide a basis for comparing student performance using school based assessments, most education authorities adopted a method of scaling the assessments across the cohort using some external instrument.

In the midst of these changes the Australian Capital Territory (ACT) educational system separated from the New South Wales (NSW) system and took the opportunity to develop a new system of senior secondary education and assessment. The main characteristics of the new system were separate year 11 and 12 schools (called colleges), no compulsory subjects and school-based curricula and assessments. However the lack of a common curriculum was problematic as content based testing couldn't be used for the scaling of school based assessments needed to put all ACT students in a subject cohort onto a single scale.

The method adopted in 1976 was to scale each subject group (or sometimes combinations of subjects when numbers were small) in each senior secondary college by their performance in a general ability test called the Australian Scholastic Aptitude Test (ASAT). Over the ensuing years the ACT education authorities formed several committees and commissioned a number of

studies to review the scaling system. These largely focused on a perception of gender bias. Although some changes were made, school-based curriculum and assessment were always held to be paramount and hence comparisons between colleges continued to be made using a general ability test. The more general problem of scaling with a psychometrically dissimilar instrument was raised in one of the studies (Masters and Beswick , 1986). In 1992 The OCS system of scaling was introduced and remains the current method. Although the OCS system introduces a third set of scores into the scaling formula it still relies on the AST as the only external instrument of comparison. These developments are discussed in detail in Chapter 1.

In recent years, several factors have made it timely to now re-investigate the use of the AST and the OCS system of scaling as the ACT's method of tertiary entrance ranking. These include increased competition for university places, concentrations of higher performing students in particular colleges and the inclusion of the AST results of culturally and linguistically diverse (CLD) students in the OCS system which had previously been excluded.

The hypothesis of this study is that the cohort a student is assessed with will have an impact on that student's scaled score and the aim is to estimate the extent of any impact found. This is done by means of a stochastic simulation of a cohort of students described in detail in chapter 5. No previous study has investigated the impact of the student's cohort in this way. Past studies have focused on suspected biases for particular groups of students. My simulation and analysis of results focuses on the correlation between the AST and course results of scaling groups because the annual statistics published by the ACT Board of Senior Secondary Studies (BSSS) indicate that they are moderate to very poor depending on the subject type. This lack of correlation suggests a significant cohort impact and invites further investigation. I therefore generate cohorts of students with authentic course/AST correlations, higher correlations and lower correlations. In each case students are removed and the difference made to the scores of the remaining students is measured.

Chapter 1 examines the historical and social context for the development of the ACT's current tertiary entrance ranking system in order to understand why a system prone to bias was originally endorsed and how it has been modified since its introduction. The events and deliberations that culminated in the original system are discussed as is the focus of previous investigations and actions taken as a result. The studies referred to by the original committee are also examined to highlight the imperatives that influenced the design of the original and current systems and to shed light on why the question of cohort impact on an individual's marks has never been investigated.

Chapter 2 discusses methods employed by other Australian jurisdictions for calculation of their Tertiary Entrance Rank. Most jurisdictions in Australia scale the school based assessment with an external device based on subject content, thus requiring a further mechanism for scaling subjects against each other. As systems of scaling based on subject content attract little controversy, other jurisdictions focus their attention on monitoring the scaling between subjects. In the ACT the use of the AST as the external scaling instrument has made the comparison of subject assessments *between schools* of greater concern. For this reason the methods used by other jurisdictions of scaling their school-based assessments within subjects but across schools is of interest to this study, especially in regards to correlations with their scaling instruments. This chapter does not seek to compare the systems of tertiary entrance ranking, but rather to provide a wider context to the ACT's system and enable the reader to understand different approaches to the scaling of school based assessment. The discussion of content based testing to scale school based assessment aims to highlight the higher correlations that tend to be achieved this way. However, the extent to which the external scaling instrument counts towards an individual's marks is beyond the scope of this thesis and therefore not considered. The governance of different jurisdictions is also examined in order to understand how the integrity of other systems is monitored and maintained.

Chapter 3 draws upon interviews and technical papers provided by the office of the ACT BSSS and explains the current ACT system in detail. Chapter 4 outlines the methodology of the simulation. It offers a rationale for the assumptions made and the decisions taken during the

construction of the model. Chapter 5 provides a comprehensive analysis and discussion of the results of the simulations.

In the conclusion I summarise my research and discuss the implications of my results.