

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Moore, Kenneth Scott

Title:

Investigating higher education productivity and its measurement in Australia

Date:

2019

Persistent Link:

<https://hdl.handle.net/11343/238674>

Terms and Conditions:

Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.

**Investigating higher education productivity and its measurement in Australia**

Kenneth Moore

ORCID ID: <https://orcid.org/0000-0002-5705-0901>

PhD Higher Education

April, 2020

University of Melbourne Graduate School of Education

Centre for the Study of Higher Education

This thesis is being submitted in total fulfilment of the degree. It is not being completed under a jointly awarded degree.

## **Abstract**

**Introduction:** This PhD dissertation focuses on higher education institutional productivity. Universities across the globe have met the 21st century with pressure to demonstrate performance and value for money. Yet common institutional performance indicators are the product of controversial choices about what data sources to consider and how data should be treated. Any individual performance indicator may be challenged on the basis of the data used to construct it. This situation raises the question: how should university performance be measured and demonstrated?

**Methods:** This thesis reviews the literature on higher education performance and productivity and documents a research program whose findings offer a novel characterisation of university productivity. Methods focus on developing productivity measures that are fit-for-purpose in context. Primary data comes from the Australian higher education system. The thesis uses a design science methodology to guide enquiry and to structure a multi-method research program that incorporates quantitative and qualitative methods. The research revolves around the development of a quantitative productivity measurement model that is iteratively tested to inform performance and productivity assessment in the Australian higher education context. Qualitative research is undertaken in the form of interviews. The interviews solicit feedback from Australian higher education experts and stakeholders. Participants provide information about the strengths and limitations of the measurement model and its results in context, as well as about opportunities for improving performance and productivity in the Australian system.

**Findings and contribution:** Findings make theoretical, technical, and practical contributions. They have implications for both policy makers and institutional leaders. Findings provide insight on relationships between teaching and research, specialisation of the academic workforce, and the role of performance measures in decision-making. Limitations include the study's intentional focus on dynamics and trends at the institution level, meaning that inter-departmental phenomena and discipline-specific phenomena are not captured. Also, due to the inconsistency of data available on service and engagement, this pillar of higher education is not included as part of academic productivity assessment.

## **Declaration**

This thesis comprises only the original work of the Author. Joint publications generated during PhD candidature and contributions of co-authors are indicated in the Preface. Due acknowledgement has been made in the text to all other documents and materials used. The thesis is fewer than the maximum word limit in length, exclusive of tables, figures, bibliographies and appendices.

Kenneth S Moore

## Preface

Seven scholarly publications were realised during this PhD program. The publications represent the work undertaken to complete this thesis. Table 1 provides a reference for each publication, the nature and contribution of co-authors and the sections of the thesis to which the publications relate. None of the publications is used as a stand-alone chapter in this thesis. The publications, rather, represent isolated components of the PhD research program. The thesis serves to synthesise all work and findings of the research and references the publications where applicable. Please note that the final two publications listed in Table 1 are book chapters within Coates (2017a) and are not directly referenced in the thesis. The full book, however, is referenced multiple times. Table 2 lists all conferences and workshops attended during candidature.

**Table 1: Scholarly publications**

| <b>Publication</b>                                                                                                                                                                               | <b>PhD Student original contribution</b>                                                                              | <b>Contributions of Co-authors</b>                                                                                      | <b>Relevant Section(s) of Thesis</b> |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------|--------------------------------------|
| <b>Moore, K.</b> , Croucher, G., Coates, H. (2019). Productivity and Policy in Higher Education. <i>Australian Economic Review</i>                                                               | 85%<br>Conceptualised study, implemented research, and authored paper.                                                | 15%<br>Provided editorial contributions, added contextual insight, refined study implications.                          | Chapters 3 & 9                       |
| <b>Moore, K.</b> , Coates, H., Croucher, G. (2019). Measuring international higher education productivity: lessons from nine countries in Asia. <i>Higher Education Forum</i> .                  | 80%<br>Refined study methods, implemented research, and authored paper.                                               | 20%<br>Developed initial conceptual and methodological frames, provided editorial contributions.                        | Chapters 4 & 6                       |
| <b>Moore, K.</b> , Coates, H., Croucher, G. (2018). Investigating applications of university productivity measurement models using Australian data. <i>Studies in Higher Education</i> .         | 90%<br>Conceptualised study, implemented research, and authored paper.                                                | 10%<br>Provided editorial contributions and refined study implications.                                                 | Chapters 6, 7 & 8                    |
| <b>Moore, K.</b> (2018). Improving higher education Productivity and its measurement: Linking productivity and student success in Australia. <i>International Journal of Chinese Education</i>   | 100%                                                                                                                  | N/A                                                                                                                     | Chapter 7                            |
| <b>Moore, K.</b> , Coates, H. & Croucher, G. (2018). Understanding and improving higher education productivity. In Hazelkorn, E., Coates, H. McCormick, A.C. (Eds.). <i>Research Handbook on</i> | 75%<br>Refined study methods, implemented research, and authored conceptual framework, methods and findings sections. | 25%<br>Developed initial conceptual and methodological frames, authored introduction, provided editorial contributions. | Chapters 3 & 5                       |

|                                                                                                                                                                                                                                                                              |                                                                                                                       |                                                                                                                             |                |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------|----------------|
| <i>Quality, Performance and Accountability in Higher Education</i> . Edward Elgar Publishing.                                                                                                                                                                                |                                                                                                                       |                                                                                                                             |                |
| <b>Moore, K.</b> , Coates, H. & Croucher, G. (2017). Understanding and improving higher education productivity. In Coates, H. (Ed.) <i>Measuring Productivity in Higher Education: Landmark insights from nine Asian countries</i> . Tokyo: Asian Productivity Organisation. | 70%<br>Refined study methods, implemented research, and authored conceptual framework, methods and findings sections. | 30%<br>Developed initial conceptual and methodological frames, authored introduction, provided editorial contributions.     | Chapters 3 & 5 |
| Coates, H. & <b>Moore, K.</b> (2017). Higher education productivity: Insights and next steps. In Coates, H. (Ed.) <i>Measuring Productivity in Higher Education: Landmark insights from nine Asian countries</i> . Tokyo: Asian Productivity Organisation.                   | 40%<br>Synthesised findings and contributed writing in multiple sections of paper.                                    | 60%<br>Developed structure of paper and key topics, authored introduction and conclusion sections as well as some findings. | Chapter 6      |

**Table 2: Conference presentations and workshops**

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>Moore, K.</b>, Croucher, G., Coates, H., Liu, L., Borden, V. (2019). Developing productivity indicators for managing higher education. Paper presentation at the EAIR 41th Annual Forum, Responsibility of Higher Education Systems: What? Why? How? Leiden University, Leiden, the Netherlands.</p> <p><b>Moore, K.</b> (2019). Measuring productivity in academia. Workshop facilitation for the Academic Workload Management Workshop, L.H. Martin Institute. Melbourne, Australia</p> <p><b>Moore, K.</b> (2019). Measuring and assessing education productivity. Invited presentation for the Australian Centre for Education Research (ACER), Melbourne, Australia.</p> <p><b>Moore, K.</b> (2019). Investigating productivity and its measurement in Australian higher education. Young Scholar Forum, Institute of Education, Tsinghua University, Beijing, China.</p> <p><b>Moore, K.</b> (2018). Reforming the assessment of higher education productivity. Symposium talk at the Conference on Higher Education Evaluation Systems for the Global Era, Tsinghua University, Beijing, China</p> | <p><b>Moore, K.</b> (2018). Measuring Productivity in Higher Education. Research findings presentation at Nous Group Technical Session. Nous Group, Melbourne, Australia.</p> <p><b>Moore, K.</b>, Coates, H. &amp; Croucher, G. (2017). Measuring higher education productivity trends: an analysis using Australian higher education key performance indicators. Paper presentation at the International Workshop for Higher Education Reform. Hiroshima, Japan.</p> <p>Coates, H., Massy, W. H., <b>Moore, K.</b> (2017). Measuring productivity in higher education. Workshop delivered for Asia Productivity Organisation (APO). Yogyakarta, Indonesia.</p> <p>Coates, H., Croucher, G. &amp; <b>Moore, K.</b> (2017). Understanding and improving higher education productivity. Paper presented at the American Educational Research Association (AERA). San Antonio, United States.</p> <p><b>Moore, K.</b> (2016) Exploring Productivity in Higher Education. Opening talk at Melbourne Centre for the Study of Higher Education Postgraduate Conference. Melbourne, Australia.</p> |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

On a final note, the quotations used to introduce each chapter have been selected from this study's interview participants. All participants are kept anonymous, but Table 27 provides some details about each participant. Further information about interview methods are provided in Chapter 7.

## Acknowledgements

I first must thank my partner and son, Eleni Aicia and Benjamin Moore, for the love, support and patience throughout this journey. Next, for the trust, encouragement and mentorship of my supervisors and advisory committee, I sincerely thank Hamish Coates, Gwilym Croucher, Leo Goedegebuure, and Ian Marshman. Your professionalism and collegiality are making important and lasting marks in the study and practice of higher education. I also thank Brenda Holt for the excellent experience while living and working at The Queen's College.

I want to acknowledge the Melbourne Centre for the Study of Higher Education for the support throughout the program and a collaboration with the Asia Productivity Organisation, which provided a platform for some of the initial work in this research project. Lastly, I must acknowledge all the cafés and libraries I frequented over the duration of the program: Cody's Café, Manor Lakes Library, Stovetop Café, Tsubu Bar, The Clyde Hotel, The Giblin-Eunson Library, The Victory State Library, Sipirok Café, Coffee Club.

Further I would like to acknowledge the scholarly outlets that valued this research and motivated me to continue progressing and publishing throughout the project (Table 1 and Table 2).

## Table of Contents

|                                                                     |           |
|---------------------------------------------------------------------|-----------|
| Abstract.....                                                       | 2         |
| Declaration.....                                                    | 3         |
| Preface.....                                                        | 4         |
| Acknowledgements.....                                               | 6         |
| Table of Contents.....                                              | 7         |
| List of Tables and Figures.....                                     | 13        |
| Terms and Abbreviations.....                                        | 15        |
| <b>Chapter 1: Introduction and Overview .....</b>                   | <b>17</b> |
| 1.1 Introduction to the Research .....                              | 18        |
| 1.1.1 Benefits and limitations of measuring productivity .....      | 18        |
| 1.1.2 Linking productivity and performance .....                    | 18        |
| 1.2.1 Research Focus .....                                          | 20        |
| 1.2.2 Research Scope.....                                           | 21        |
| 1.3.1 Rationale for the Study .....                                 | 22        |
| 1.3.2 Contributions to Knowledge and Practice .....                 | 23        |
| 1.4.1 Overview of Methods .....                                     | 24        |
| 1.4.2 Methodology and Research Design.....                          | 25        |
| 1.4.3 Chapter Structure .....                                       | 25        |
| 1.5 Summary .....                                                   | 26        |
| 1.5.1 Focus and Contributions.....                                  | 26        |
| 1.5.2 Considerations and Limitations Arising.....                   | 27        |
| <b>PART I:.....</b>                                                 | <b>28</b> |
| <b>Chapter 2: The Context of Higher Education Productivity.....</b> | <b>29</b> |
| 2.1 Introduction.....                                               | 30        |
| 2.2 What is Higher Education Productivity? .....                    | 31        |
| 2.2.1 No Single Agreed Definition of Productivity.....              | 31        |
| 2.2.2 Functional Framing.....                                       | 31        |
| 2.2.3 Systematic Framing .....                                      | 32        |
| 2.3 Why is Productivity Important?.....                             | 34        |
| 2.3.1 Public Sector Accountability .....                            | 34        |

|                                                                           |           |
|---------------------------------------------------------------------------|-----------|
| 2.3.2 Emergence and Evolution of Trends across the Sector .....           | 35        |
| 2.4 Measurement Remains a Challenge .....                                 | 40        |
| 2.4.1 Opposing Views on Measurement .....                                 | 40        |
| 2.4.2 Politics, Power and Trust .....                                     | 41        |
| 2.4.3 Overcoming Challenges .....                                         | 42        |
| 2.5 How to Address Quality .....                                          | 43        |
| 2.5.1 Treatment of Quality .....                                          | 43        |
| 2.5.1 Capturing and Tracking Quality .....                                | 44        |
| 2.5.2 Quality, Productivity and Value .....                               | 44        |
| 2.6 Why and How to Improve Measurement .....                              | 45        |
| 2.6.1 Reasons to Improve .....                                            | 45        |
| 2.6.2 Measured Results Should Align with Observable Behaviour .....       | 46        |
| 2.6.3 Productivity is a Lead Indicator for Longer-term Achievement .....  | 47        |
| 2.6.4 Analytical Framings of Productivity Need Refining .....             | 47        |
| 2.6.5 Alternative Metrics are Needed at the Institution-Level .....       | 48        |
| 2.7 Summary .....                                                         | 49        |
| 2.7.1 Focus and Contributions .....                                       | 49        |
| 2.7.2 Considerations and Limitations Arising .....                        | 50        |
| <b>Chapter 3: Concepts and Technique of Productivity Measurement.....</b> | <b>51</b> |
| 3.1 Introduction .....                                                    | 52        |
| 3.2 Production Theory .....                                               | 52        |
| 3.2.1 First Principles .....                                              | 52        |
| 3.2.2 The Production Function .....                                       | 55        |
| 3.2.3 The Production Frontier .....                                       | 57        |
| 3.2.4 Productivity Change .....                                           | 57        |
| 3.3 Productivity and Education .....                                      | 59        |
| 3.3.1 Departure from Foundational Economic Assumptions .....              | 59        |
| 3.3.2 Human Capital Theory and Educational Rates of Return .....          | 60        |
| 3.3.4 Toward Institutional Production Functions .....                     | 61        |
| 3.4 Taxonomy of Productivity Measurement Models .....                     | 62        |
| 3.4.1 A Universe of Measurement Models .....                              | 62        |
| 3.4.2 Input and Output Characteristics .....                              | 63        |
| 3.4.4 Production Complexity .....                                         | 64        |
| 3.4.5 Data Aggregation Method .....                                       | 65        |
| 3.5 Contemporary Models and Gaps in Knowledge .....                       | 67        |

|                                                            |           |
|------------------------------------------------------------|-----------|
| 3.5.1 <i>Ad Hoc</i> Nature of Current Approaches.....      | 67        |
| 3.5.2 Specifying Gaps and Opportunities .....              | 69        |
| 3.6 Research Question Formulation.....                     | 71        |
| 3.7 Summary .....                                          | 74        |
| 3.7.1 Focus and Contributions.....                         | 74        |
| 3.7.2 Considerations and Limitations Arising .....         | 74        |
| <b>Chapter 4: Research Framework and Methodology .....</b> | <b>76</b> |
| 4.1 Introduction.....                                      | 77        |
| 4.2 Research Methodology .....                             | 77        |
| 4.2.1 Design Science Research.....                         | 77        |
| 4.2.2 Principles .....                                     | 78        |
| 4.2.3 Practice .....                                       | 79        |
| 4.2.4 Procedures .....                                     | 79        |
| 4.3 Research Design and Methods.....                       | 80        |
| 4.3.1 Design and Framework.....                            | 80        |
| 4.3.2 Methods .....                                        | 81        |
| 4.4 Summary .....                                          | 83        |
| 4.4.1 Focus and Contributions.....                         | 83        |
| 4.4.2 Considerations and Limitations Arising .....         | 84        |
| <b>PART II: .....</b>                                      | <b>85</b> |
| <b>Chapter 5: Initial Measurement Model Design.....</b>    | <b>86</b> |
| 5.1 Introduction.....                                      | 87        |
| 5.2 Initial Design Decisions.....                          | 88        |
| 5.2.1 Measurement Opportunities .....                      | 88        |
| 5.2.2 Contextual Opportunities.....                        | 88        |
| 5.2.3 Model Usage Guidelines .....                         | 90        |
| 5.3 Model Constructs.....                                  | 90        |
| 5.3.1 Constructs Overview .....                            | 90        |
| 5.3.1 Input Constructs.....                                | 91        |
| 5.3.2 Output Constructs .....                              | 93        |
| 5.4 Construct Associations.....                            | 94        |
| 5.4.1 Construct Associations Overview .....                | 94        |
| 5.4.2 Input Associations .....                             | 94        |
| 5.4.3 Output Associations.....                             | 95        |
| 5.5 Summary .....                                          | 99        |

|                                                                |            |
|----------------------------------------------------------------|------------|
| 5.5.1 Focus and Contributions .....                            | 99         |
| 5.5.2 Considerations and Limitations Arising .....             | 99         |
| <b>Chapter 6: Formative Evaluation and Model Redesign.....</b> | <b>101</b> |
| 6.1 Introduction.....                                          | 102        |
| 6.2 Asia-Pacific Model Evaluation .....                        | 103        |
| 6.2.1 Evaluation Methods.....                                  | 103        |
| 6.2.2 Model Constructs Assessment.....                         | 104        |
| 6.2.3 Construct Associations Assessment .....                  | 106        |
| 6.2.4 Model Boundaries Assessment.....                         | 108        |
| 6.2.5 Revelation, Novelty, Importance Assessment.....          | 111        |
| 6.2.6 Asia Pacific Evaluation Conclusions.....                 | 113        |
| 6.3 Australian Data Evaluation .....                           | 114        |
| 6.3.1 Evaluation Methods and Measurement Results.....          | 114        |
| 6.3.2 Model Constructs Assessment.....                         | 116        |
| 6.3.3 Construct Associations Assessment .....                  | 117        |
| 6.3.4 Model Boundaries Assessment.....                         | 118        |
| 6.3.5 Revelation, Novelty, Importance Assessment.....          | 119        |
| 6.3.6 Australian Data Evaluation Conclusions.....              | 120        |
| 6.4 Australian Model Instantiation .....                       | 121        |
| 6.4.1 Model Redesign.....                                      | 121        |
| 6.4.2 Modelling Environment Change .....                       | 123        |
| 6.4.3 Capital Service Estimation .....                         | 123        |
| 6.4.4 Input Smoothing .....                                    | 124        |
| 6.4.5 Output Value Weight Ranges .....                         | 126        |
| 6.4.6 Cumulative Productivity Change.....                      | 127        |
| 6.4.7 Productivity Threshold Estimation.....                   | 128        |
| 6.5 Model Re-design Results .....                              | 128        |
| 6.6 Summary .....                                              | 133        |
| 6.6.1 Focus and Contributions.....                             | 133        |
| 6.6.2 Considerations and Limitations Arising .....             | 134        |
| <b>Chapter 7: Formal Evaluation and Model Enhancement.....</b> | <b>136</b> |
| 7.1 Introduction.....                                          | 137        |
| 7.2 Evaluation Methods .....                                   | 138        |
| 7.2.1 Methodology.....                                         | 138        |
| 7.2.2 Participants .....                                       | 139        |

|                                                                        |            |
|------------------------------------------------------------------------|------------|
| 7.2.3 Data Collection and Analysis .....                               | 140        |
| 7.3 Expert and Stakeholder Evaluation.....                             | 141        |
| 7.3.1 Underlying Constructs Evaluation.....                            | 141        |
| 7.3.2 Model Constructs Assessment.....                                 | 145        |
| 7.3.3 Construct Associations Assessment .....                          | 147        |
| 7.3.4 Model Boundaries Assessment.....                                 | 148        |
| 7.3.5 Revelation, Novelty, Importance Assessment.....                  | 149        |
| 7.3.6 Evaluation Conclusions .....                                     | 153        |
| 7.4 Australian Model Enhancement.....                                  | 154        |
| 7.4.1 Model Redesign.....                                              | 154        |
| 7.4.2 Real Value for Money Testing .....                               | 155        |
| 7.4.3 Productivity Drivers Calculation .....                           | 156        |
| 7.4.4 Single Factor Productivity Labour Inputs .....                   | 156        |
| 7.4.5 Separate Productivity Calculation .....                          | 157        |
| 7.4.6 Education and Research Output Consistency .....                  | 158        |
| 7.4.7 Approximating Absolute Productivity with ‘Relative Change’ ..... | 159        |
| 7.6 Summary .....                                                      | 162        |
| 7.6.1 Focus and Contributions.....                                     | 162        |
| 7.6.2 Considerations and Limitations Arising .....                     | 163        |
| <b>PART III:.....</b>                                                  | <b>164</b> |
| <b>Chapter 8: Measurement Results and Productivity Assessment.....</b> | <b>165</b> |
| 8.1 Introduction.....                                                  | 166        |
| 8.2 TFP Change vs. SFP Change .....                                    | 168        |
| 8.2.1 TFP and Change Thresholds.....                                   | 168        |
| 8.2.2 SFP Change and Thresholds.....                                   | 173        |
| 8.3 Separate SFP Change.....                                           | 176        |
| 8.4 Absolute TFP Change.....                                           | 179        |
| 8.5 Emergent Patterns and Deeper Dynamics.....                         | 182        |
| 8.5.1 Emergent Patterns.....                                           | 182        |
| 8.5.2 Dynamics of Data Elements .....                                  | 183        |
| 8.5.3 Model Versions Insights.....                                     | 186        |
| 8.6 Summary .....                                                      | 189        |
| 8.6.1 Focus and Contributions.....                                     | 189        |
| 8.6.2 Considerations and Limitations Arising .....                     | 189        |
| <b>Chapter Nine: Discussion of Findings and Implications .....</b>     | <b>191</b> |

|                                                               |     |
|---------------------------------------------------------------|-----|
| 9.1 Research Summary .....                                    | 192 |
| 9.1.1 The Value of the Research.....                          | 192 |
| 9.1.2 Accumulated Learnings.....                              | 192 |
| 9.2 Findings and Implications.....                            | 193 |
| 9.2.1 Presentation of Findings .....                          | 193 |
| 9.2.2 Addressing Research Question 1 .....                    | 194 |
| 9.2.3 Addressing Research Question 2.....                     | 197 |
| 9.2.4 Addressing Research Question 3.....                     | 201 |
| 9.2.5 Implications for Policy and Practice.....               | 205 |
| 9.3 Limitations and Future Research .....                     | 212 |
| 9.3.1 Incorporating Quality .....                             | 212 |
| 9.3.2 Measuring Individual Academic Functions.....            | 213 |
| 9.3.3 Interdepartmental Dynamics.....                         | 213 |
| 9.3.4 Productivity Measurement and Assessment Techniques..... | 214 |
| 9.4 Conclusions and Consequences .....                        | 214 |
| 9.4.1 What has been Learned? .....                            | 214 |
| 9.4.2 What are the Consequences? .....                        | 215 |
| References.....                                               | 217 |
| Appendix A.....                                               | 230 |
| Appendix B .....                                              | 231 |
| Appendix C .....                                              | 232 |
| Appendix D.....                                               | 235 |
| Appendix E .....                                              | 238 |
| Appendix F.....                                               | 239 |
| Appendix G.....                                               | 240 |
| Appendix H.....                                               | 242 |
| Appendix I .....                                              | 244 |

## List of Tables and Figures

|                                                                                                        |     |
|--------------------------------------------------------------------------------------------------------|-----|
| Table 1: Scholarly publications .....                                                                  | 4   |
| Table 2: Conference presentations and workshops.....                                                   | 5   |
| Table 3: Scope of Research.....                                                                        | 21  |
| Table 4: Clarification of terms.....                                                                   | 22  |
| Table 5: Productivity models in higher education .....                                                 | 68  |
| Table 6: Initial prototype measurement model specifications .....                                      | 87  |
| Table 7: Input constructs and specifications .....                                                     | 91  |
| Table 8: Output constructs and specifications .....                                                    | 93  |
| Table 9: Data collection by each participating country (Moore, Coates, et al., 2019). .....            | 103 |
| Table 10: Construct assessment findings.....                                                           | 104 |
| Table 11: Country-by-country output representation decisions.....                                      | 105 |
| Table 12: Construct associations assessment findings.....                                              | 106 |
| Table 13: Pakistan source data (Coates, 2017, p. 168) .....                                            | 107 |
| Table 14: Model boundaries assessment findings .....                                                   | 108 |
| Table 15: Revelation, novelty and importance assessment findings .....                                 | 111 |
| Table 16: Adapted NRC model input-output framework for Australian data .....                           | 115 |
| Table 17: Adapted NRC model results on Australian data.....                                            | 115 |
| Table 18: Construct evaluation findings .....                                                          | 116 |
| Table 19: Construct associations assessment findings.....                                              | 117 |
| Table 20: Model boundaries assessment findings .....                                                   | 118 |
| Table 21: Revelation, novelty and importance assessment findings .....                                 | 120 |
| Table 22: Redesigned model input-output framework .....                                                | 122 |
| Table 23: Pakistan reported productivity change indexes and corresponding $\Pi P\Delta t$ values. .... | 128 |
| Table 24: Redesigned model productivity change and cumulative change results .....                     | 129 |
| Table 25: Cumulative change comparison for different models 2007-2013 .....                            | 130 |
| Table 26: Redesigned model rankings for different value weight schemes 2007-2013.....                  | 132 |
| Table 27: Participant Codes.....                                                                       | 139 |
| Table 28: Different classifications of productivity paradigms in recent literature .....               | 142 |
| Table 29: Productivity paradigms identified in the current research .....                              | 143 |
| Table 30: Model construct assessment findings .....                                                    | 146 |
| Table 31: Construct associations assessment findings.....                                              | 147 |
| Table 32: Model boundaries assessment findings .....                                                   | 148 |
| Table 33: Revelation, novelty and importance assessment findings .....                                 | 149 |
| Table 34: Participant-identified areas of inefficiency for universities.....                           | 151 |
| Table 35: Model input-output framework.....                                                            | 154 |
| Table 36: Summary of all essential model design and re-design innovations.....                         | 161 |
| Table 37: Key variables and system index values 2007-2016.....                                         | 167 |
| Table 38: Institutional rankings' sensitivity to different value weight schemes .....                  | 172 |
| Table 39: System Drivers of 50-50 TFP and SFP Change 2007-2016.....                                    | 183 |
| Table 40: Diversity of institutional TFP growth trends 2007-2016.....                                  | 184 |

|                                                                                           |     |
|-------------------------------------------------------------------------------------------|-----|
| Table 41: Classifying productivity change trends .....                                    | 185 |
| Table 42: Productivity measurement model design principles .....                          | 197 |
| Table 43: Constraints to productivity improvement .....                                   | 211 |
| Table 44: Measurement and assessment limitations.....                                     | 214 |
| Table 45: Input-output framework for APO model test.....                                  | 230 |
| Table 46: Institutional absolute productivity indexes 2007 .....                          | 242 |
| Table 47: Model version comparison, system TFP change results for each year 2007-13 ...   | 244 |
| Table 48: Model version comparison, cumulative change results 2007-13.....                | 244 |
| Table 49: Real value for money adjustment compared to Moradi-Motlagh et al. (2016) .....  | 245 |
| Table 50: Institutional rankings' sensitivity to different value weight schemes .....     | 246 |
|                                                                                           |     |
| Figure 1: Research framework.....                                                         | 25  |
| Figure 2: Higher education performance system components (Miller, 2007, p. 38) .....      | 33  |
| Figure 3: Higher education institution performance system (Miller, 2007, p. 132) .....    | 33  |
| Figure 4: Input-output performance spectrum .....                                         | 56  |
| Figure 5: Single input and output production frontier curve .....                         | 57  |
| Figure 6: Production function taxonomy .....                                              | 62  |
| Figure 7: Productivity models used in this research .....                                 | 70  |
| Figure 8: Space for scholarly contribution.....                                           | 71  |
| Figure 9: Design Science Research Cycles (Hevner, 2007) .....                             | 79  |
| Figure 10: Eight-stage research project framework.....                                    | 81  |
| Figure 11: Research Stage 2 .....                                                         | 87  |
| Figure 12: Model construct associations and data aggregation schematic .....              | 98  |
| Figure 13: Research Stages 3 & 4.....                                                     | 102 |
| Figure 14: Reported change figures from Pakistan (Coates, 2017, p. 168) .....             | 107 |
| Figure 15: Fluctuations in Pakistan's source data.....                                    | 110 |
| Figure 16: Fluctuations in Cambodia's component input indexes .....                       | 110 |
| Figure 17: Indonesia education, research and composite indexes (Coates, 2017a, p. 106)... | 111 |
| Figure 18: Spikes in measured yearly productivity change .....                            | 119 |
| Figure 19: Measurement model schematic with highlighted elements for revision .....       | 121 |
| Figure 20: Model redesign schematic .....                                                 | 123 |
| Figure 21: $IP\Delta t$ thresholds for four universities .....                            | 133 |
| Figure 22: Research Stages 5 & 6.....                                                     | 137 |
| Figure 23: Design science relevance cycle .....                                           | 139 |
| Figure 24: Final model iteration schematic .....                                          | 155 |
| Figure 25: Research Stage 7 .....                                                         | 166 |
| Figure 26: System TFP thresholds 2007-2016 .....                                          | 168 |
| Figure 27: Institutional TFP change thresholds .....                                      | 170 |
| Figure 28: System SFP change thresholds 2007-2016 .....                                   | 174 |
| Figure 29: Joint SFP change threshold estimates .....                                     | 175 |
| Figure 30: Separate SFP change system trends 2007-2016 .....                              | 177 |
| Figure 31: Separate productivity change estimates per institution .....                   | 178 |
| Figure 32: Absolute TFP approximation and relative change estimates .....                 | 181 |
| Figure 33: Comparison of productivity TI and DEA estimates for three select universities. | 187 |
| Figure 34: Research Stage 8 .....                                                         | 192 |
| Figure 35: Approaches to understanding higher education productivity.....                 | 194 |

## Terms and Abbreviations

|              |                                                                   |
|--------------|-------------------------------------------------------------------|
| ABS          | Australian Bureau of Statistics                                   |
| APC          | Australia Productivity Commission                                 |
| APO          | Asia Productivity Organization                                    |
| Assessment   | Using standards and evidence to evaluate some topic or phenomenon |
| BLS          | Bureau of Labor Statistics                                        |
| CGS          | Commonwealth Grant Scheme                                         |
| CPI          | Consumer price index                                              |
| DEA          | Data envelopment analysis                                         |
| DET          | Department of Education and Training                              |
| EBA          | Enterprise bargaining agreement                                   |
| EFTSL        | Equivalent full-time student load                                 |
| Go8          | Group of Eight                                                    |
| HEI          | Higher education institution                                      |
| HESA         | Higher Education Support Act                                      |
| Indicator    | Numeric proxy representing a difficult-to-measure phenomenon      |
| KS           | Key stakeholder                                                   |
| Measure      | Numeric value resulting from the act of measurement               |
| Measurement  | Assigning a number to some object, event or phenomenon            |
| Metric       | Computed value or ratio derived from one or more measure          |
| NRC          | National Research Council                                         |
| OECD         | Organisation for Economic Cooperation and Development             |
| PCS          | Provider Category Standards                                       |
| Performance  | Level of achievement in some domain                               |
| Productivity | Output per unit input                                             |
| SFA          | Stochastic frontier analysis                                      |
| TEQSA        | Tertiary Education Quality Standards Association                  |
| TE           | Technical expert                                                  |
| TI           | Törnqvist index                                                   |



## CHAPTER 1: INTRODUCTION AND OVERVIEW

*Productivity is a very idiosyncratic thing.*

–Study participant TE10,  
Former University Chief Financial Officer

### **Abstract:**

This chapter provides an overview of this dissertation. It summarises major concepts and contributions, as well as the study's scope and structure. The chapter first introduces the topic of productivity as it relates to higher education. It then covers the guiding research questions of the project to frame the scope for inquiry. The methodological frame of design science is then introduced to explain how the project addresses the research questions. Finally, deliberate limitations in the scope for inquiry are acknowledged, and a summary of each chapter is provided.

### **Key Words:**

Higher education, productivity, measurement, assessment, policy, practice

## **1.1 Introduction to the Research**

### **1.1.1 Benefits and limitations of measuring productivity**

This research investigates productivity and performance measurement in higher education. It uncovers empirical trends and exposes inroads for improving institutional and system-wide assessment and evaluation. Research insights further demonstrate the importance of assessing productivity both to reflect and to inform institutional behaviour and policy.

The research explores the measurement of university productivity at the institutional level. Insights derive largely from examination of quantitative performance indicators. Numerical and mathematical representations of university performance, however, will have limitations. Any single choice for measurement brings inevitable trade-offs regarding alternative institutional phenomena that could be represented in analysis. Measurement limitations, however, should not overshadow the benefits of observing and learning from empirical trends. This research attempts to maximise the insight of quantitative institutional analysis by incorporating rigorous qualitative study and expert judgement. The research thus progresses with both ambition and restraint for how much information quantitative institutional performance and productivity indicators can capture and represent.

Productivity is understood as one aspect of institutional performance. This dissertation positions productivity measurement as a useful tool for higher education performance assessment. The study explores technical and mathematical foundations of higher education productivity, but it does not advance a positivist, context-free interpretation of quantitative findings. The research seeks to extract as much value from university data as possible, but no more. Rather than targeting the utmost precision in measured results, this research champions techniques that embrace variability and uncertainty in higher education institutional performance data. Findings are both quantitative and qualitative, and analysis is introspective regarding the practice of productivity measurement in higher education. Multiple experts and stakeholders are involved in calibrating methods and expectations for assessment. Resultant findings provide novel insight on aspects of higher education policy and provision. Specific implications are discussed for the Australian higher education system.

### **1.1.2 Linking productivity and performance**

Trends and data considered for institutional performance assessment should be able to reflect authentic performance and behaviour. Performance information should further be able to underpin rationales behind performance improvement initiatives and policy. If executed well, the results of performance assessments should be able to inform the following questions: *how should a university function to deliver the promise of its mission, and how should institutional leaders use new information for real-time decision-making?*

Diverse stances on purposes and functions of higher education, however, leave no simple or single solution to problems concerning the extent to which data analysis should be used to

design institutional strategies, policies and interventions. With diverse and changing ideas about universities' place in society and the scope of academic work, objective information is needed, but expert judgement on interpretation of data is also essential. This dissertation does not seek to validate any single operational model or strategy. Rather, it investigates different ways for generating useful information on how institutions function. It exposes evidence that merits consideration, not only in terms of how past policy decisions have likely driven key trends, but also how measured trends may inform future decision-making.

The role of performance measurement in higher education is important. The numbers produced by measurement, however, may not provide as much insight as reflecting on how the measurement exercise was conducted. A persistent challenge for measuring productivity is the myriad colloquial uses and understandings of the term. Productivity is a loaded concept. Different stakeholders conceptualise the term in different ways. This dissertation demonstrates that measurement practice must reflect this reality. Many common, contemporary techniques for measuring university productivity are too monolithic and impenetrable to offer actionable insight. While this dissertation argues for the importance of valid and reliable measures of productivity, it simultaneously demonstrates trade-offs between increasing the precision of a performance metric and maintaining its usefulness.

The research methods and the nature of results in this dissertation stand in contrast to prior work in the field. This project commences with a first-principles examination of productivity measurement for university performance assessment. Current literature on higher education productivity frequently employs methods imported from other fields. The bulk of research focuses on the iterative optimisation of well-developed quantitative models to increase statistical reliability and robustness in relation to particular datasets (De Witte & Sneyers, 2017). This dissertation, however, develops a context-sensitive measurement technique and focuses on alternative understandings of data. Most studies in the literature offer a singular interpretation of productivity. They offer a rationale for a single, selected measurement technique and present ranked findings in terms of better and worse performance. The current study uses variations in the treatment of data to address multiple interests in university performance. It then presents findings in terms of emergent patterns and differences that unfold with successive portrayals of data. It does not presume to make final, conclusive statements about single institutions' performance based on a "best fit" model.

### **1.1.3 Defining Productivity**

This section offers operative, high-level definitions and understandings of productivity, efficiency and effectiveness for a shared understanding of the terms throughout the dissertation. First, we examine the productivity of institutions, not of individual people. Productivity is understood as a comparison of institutional outputs to inputs. In a mathematical sense, productivity is the ratio of quantity output per unit input, or  $O/I$ . This understanding requires an assumption that higher education institutions create value by performing functions that take some measurable input and transform it into some measurable output (Gordon & Vaughan, 2011).

Tangen (2005) discusses the importance of distinguishing between mathematical definitions of productivity and practical understandings. In a practical sense, this dissertation defines productivity in terms of efficiency and effectiveness. Productivity represents both the

efficiency and the effectiveness of key institutional work processes. Efficiency revolves around minimising that time and resources needed to accomplish a set of given objectives. Effectiveness revolves around maximising the creation of value given clear resource constraints (Sink & Tuttle, 1989; Sumanth, 1994).

The academic literature often conflates efficiency and productivity. This dissertation ascribes them different meanings. Efficiency refers only to the input-centric component of productivity. Hence, this research never assumes that an increase in efficiency automatically means an increase in productivity. For example, efficiency at the expense of effectiveness is not considered a productivity gain.

Effectiveness, in the literature, is even more diffuse than productivity and efficiency. It is the most difficult to quantify objectively (Tangen, 2005), which is why productivity measurement in higher education needs careful consideration. Chapters 2 and 3 further explore definitions and understandings of these terms as they relate to higher education.

## **1.2 Focus and Scope**

### **1.2.1 Research Focus**

The focus of the study is to investigate and to position productivity measurement as a viable assessment mechanism for university performance. Chief aims of the project include the demonstration and characterisation of the limits and the potential of higher education productivity measurement. An objective includes parameterising the use of productivity measurement results to generate novel, valuable and actionable conclusions for informing university performance improvement. Most of the analysis centres on the Australian context, but the insight generated transcends national boundaries. The research explores and develops productivity measurement in general. It targets indexing measurement methods for estimating productivity change, focusing on the Törnqvist index (TI). The TI was advanced by Caves, Christensen, & Diewert (1982) and introduced for use in higher education by Sullivan, Mackie, Massy, & Sinha (2012).

The focus of the research is further clarified via this study's guiding research question.

- Q1.** How can different approaches to understanding productivity in higher education be characterised?
- Q2.** What are the characteristics of a measurement model appropriate for investigating productivity of Australian universities?
- Q3.** How can the productivity of Australian universities be characterised?

The questions allow for concurrent investigation of the conceptual, technical and practical underpinnings of productivity measurement for performance assessment in higher education. The first question regards documenting and explaining of a range of theoretical and contextual understandings of productivity. The second question regards the characteristics of a measurement model that reflect both theoretical and contextual considerations for

appropriate analysis of Australian universities. The third question targets the shape and nature of the estimates such a model could produce for the Australian system. Using results from addressing the questions, it is then important to demonstrate how findings could form the basis of a performance assessment and illuminate implications for policy, practice and decision-making.

### 1.2.2 Research Scope

In measuring and investigating university productivity, this study limits the scope of inquiry in four key ways. First, this study examines only the research and education functions of universities. Second, it generates only institution and system-level indicators. Third, it assumes that quality is inherent to productivity, and finally it measures productivity change, as opposed to absolute productivity. Details about the scope are explained in Table 3.

**Table 3: Scope of Research**

| <b>Parameter</b>            | <b>Inclusion</b>                  | <b>Exclusion</b>                  | <b>Rationale</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|-----------------------------|-----------------------------------|-----------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| University Function         | Education and Research            | Service and Engagement            | Exclusion of service and engagement is not intended to signal any relative importance of this university function. In the Australian case, there exist no system-wide, comprehensive definitions of this function nor set parameters for what minimum output constitutes an achievement in this domain. Defining a new, subjective measure of service and engagement would likely confound results as much as add explanatory value. Both prior literature and intermediate research findings support the limited scope of measuring only education and research.   |
| Level of Analysis           | System and Institution            | Individual and Inter-departmental | The decision to exclude interdepartmental or discipline-related dynamics also relates to source data availability and reliability. Department-level dynamics have important value in explaining what is driving institution-level trends, but accessing data in a consistent manner across Australian institutions would require full institutional case studies, which goes beyond the scope of this project.                                                                                                                                                      |
| Performance Dimension       | Inputs, Outputs, and Productivity | Quality and Outcomes              | Quality is acknowledged as crucial to performance and productivity. Explicit quality indicators, however, are excluded because of subjectivity and reliability. It is assumed in this study that quality must be part of productivity. That is, a measured productivity improvement should not be considered a real productivity improvement if it is at the expense quality. At this stage, however, parallel assessments of both productivity and quality are required to supplement the findings of the other, and the current research focuses on productivity. |
| Productivity Characteristic | Productivity Change               | Absolute Productivity             | Comparing the absolute performance or productivity of institutions in different contexts with different missions is contentious at the outset. This study thus focuses on using time-series data to understand trends over time and institutional performance trajectories, rather than making statements about absolute levels of performance.                                                                                                                                                                                                                     |

Defining further terminology is also important in clarifying the scope of research. First, the terms ‘measurement’ and ‘assessment’ are understood as different processes with different objectives. Relevant and associated terminology are also listed and defined below. The terms

listed in Table 4 are defined inconsistently across higher education literature (Borden & Bottrill, 1994). Explicit definitions are thus provided to establish a consistent understanding and usage with respect to the scope of the current project.

**Table 4: Clarification of terms**

| <b>Term</b> | <b>Definition and usage</b>                                                                                                                                                 |
|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Measurement | Measurement is the assignment of numbers to aspects and qualities of objects and events according to some rule or convention (Pedhazur & Schmelkin, 1991).                  |
| Assessment  | Assessment is the collecting, analysing, interpreting and using of information for the purpose of improving performance, relative to standards or objectives (Paige, 2005). |
| Measure     | A measure is a number attained from the act of measurement, as per the definition above.                                                                                    |
| Metric      | A metric refers to a composite value derived from one or more measures and usually represents a more abstract concept than that of a measure (NIST, 2008).                  |
| Indicator   | An indicator is a substitute for a precise measure of some phenomenon that is difficult to quantify (Paige, 2005).                                                          |

The operative definition of ‘measurement’ is left intentionally broad because the chapters that follow cite a substantial body literature on the topic. The broad definition captures the range of usage of the term across different contexts. Because productivity is inherently a comparison of inputs and outputs, a productivity measure may also be referred to as a metric or as an indicator. Productivity, though, is most commonly referenced as a self-contained phenomenon, which is emergent from its sub-components. This research focuses on measuring events of productivity change over time.

The term ‘higher education’ also merits definition. In this dissertation, it is primarily used as a reference to universities, where universities are institutions that engage in both teaching and research and offer post-graduate-level qualifications. Higher education may be distinguished from ‘tertiary’ or ‘postsecondary’ education. Tertiary and post-secondary refer to a broader spectrum of institutions that offer accredited professional, vocational and higher degrees after formal schooling. In Chapter 2, higher education is also used more abstractly to capture broader ideas about the scope and nature of academic work.

## **1.3 Contributions and Impact**

### **1.3.1 Rationale for the Study**

The benefits of enhancing data-driven understandings of institutional performance outweigh concerns about the misuse of individual metrics. Even if quantitative performance portrayals are limited, neglecting available and reliable data represents only a missed opportunity to learn from the efforts and the resources spent to collect, store, and report information about

universities. A substantial literature pits measurement against quality and against the integrity of academic work (Feller, 2009; Kallio & Kallio, 2014; Muller, 2018; Olssen & Peters, 2005). While this dissertation acknowledges that performance measurement will always have some dimension of politics and power (Dooren & Van de Walle, 2008), the primary framing of productivity measurement in this dissertation is one of empirical science and institutional research—not a framing of power, control, and trust. Better productivity indicators are necessary for better understanding and for improving productivity. To improve productivity measurement is to become more adept at identifying and articulating what is important to improve.

This research is sympathetic, however, to the incentive structures that performance measures can either create or reinforce. International rankings represent a prime example of how institution-level performance metrics can muddy the waters of institutional strategy and decisions regarding resource allocation. If it is the case that institutional rankings create perverse incentives for institutions (Coates, 2016; Dearden, Grewal, & Lilien, 2008; Hazelkorn, 2015; Marginson & van der Wende, 2007), it is the position of this dissertation that those incentives are best combatted by more and better information, not less. When institution-level performance indicators are not being used for rankings, they are being used for accreditation activities. This has led Astin (2012) to the conclusion that practical and operational use of measurement has been largely neglected within institutions. Determining and characterising new measurement and assessment approaches for practical purposes is a key objective of this study.

### **1.3.2 Contributions to Knowledge and Practice**

Implications for higher education policy and provision constitute prime contributions of this dissertation. First, quantitative results illustrate ranges of institutional performance portrayals based on key performance indicators. The demonstration of alternative interpretations of data signal risks to system-level performance funding policy. Findings further provide evidence for both institutional differentiation and trends toward isomorphism. Discussion targets the drivers and consequences of instituting policy concerning the segmentation of institutions within a system. Information transparency and reporting is also discussed. The risks and benefits of financial transparency are weighed. Issues of practice regard benefits and trade-offs of institutions pursuing a productivity improvement agenda, and common constraints to productivity improvement are identified.

The research further identifies and describes different stakeholder understandings of higher education productivity. Differing views on the topic are discussed in relation to data interpretation and stakeholder arbitration in decision-making contexts. The dissertation demonstrates how varied interests in data and performance can be acknowledged and addressed in measurement. Findings also help to identify generalisable design principles for developing higher education productivity measurement models. Productivity measurement in higher education must be technically and conceptually sound. It must also be fit-for-purpose in the environment where it is intended to be used. The design principles identified diverge from conventional knowledge and practice on higher education productivity measurement. They address implicit but impactful implications of how data are selected and treated in a productivity measurement model.

This dissertation shows that measurement transparency is essential for the development and practical use of institution-level metrics and indicators. The research demonstrates, however, transparency in the treatment of data cannot be handled casually. In contrast to most research in the field, this dissertation emphasises learnings that derive from the exercise of measurement over learnings that derive from precise magnitudes of measured results. Conventional university productivity studies have gravitated toward the application of advanced mathematical methods to calculate productivity ratios. Such techniques allow for the analysis of multi-variate data and require no explicit input or judgement from researchers about the value of inputs and outputs (Carrington, Coelli, & Rao, 2005; Duan, 2019). This dissertation argues that considering values and interests, as well as examining the drivers of observed trends, is what establishes the utility of a performance measurement exercise—more so than prioritising the clinical precision of measurement results.

Simple measurement techniques may also be preferable to more complex methods. There is value in employing simpler mathematical methods that require more intervention and intermediate decision-making by researchers. It showcases the TI technique by adapting a model first developed for higher education by the National Research Council (NRC) (Sullivan et al., 2012). TI methods are less attractive in the field because of the absence of off-the-shelf software applications that employ the technique. Manual application of TI methods, however, increase a researcher's ability to decompose measured results and offer control over how data elements are emphasised in calculation. That is, do data aggregation tasks are relegating to complex computer algorithms. On one hand, TI measurement results that rely on researcher judgements are contestable. On the other hand, by bringing quantitative assumptions to the forefront—in a manner easily communicated to non-experts—TI methods likewise demonstrate how any single measurement of institutional productivity is contestable. No matter the calculation technique, when measuring institution-level productivity and performance, this dissertation shows the extent to which the field needs improved protocol for increasing relevance and suitability in context, rather than statistically optimised mathematical models (De Witte & Sneyers, 2017).

## **1.4 Project Design and Dissertation Structure**

### **1.4.1 Overview of Methods**

Key methods in this research include large-scale collection of secondary institutional performance data on 38 Australian universities. The data is available from public sources and covers the period 2007-2016. The research proceeds to design and develop a TI measurement model for estimating institutional productivity change. The model is tested across ten countries' data in the Asia-Pacific, including Australia, to reveal broad strengths and limitations of the model. Structured interviews are then conducted with 20 Australian experts and stakeholders to influence continued development and adaptation of the model for the Australian context. Interview responses help to expose and clarify Australia-specific higher education productivity issues and help to inform context-dependent interpretation of model

results. The project concludes with a final comprehensive data analysis and a productivity assessment of Australian universities using the data produced by the model and using responses from the interviews.

### 1.4.2 Methodology and Research Design

All methods were selected and structured according to design science research methodology. Design science produces new knowledge and practical solutions through the design of some ‘artefact’ (Peppers, Tuunanen, Rothenberger, & Chatterjee, 2007). The artefact in this research is a productivity measurement model. The methodology incorporates three concurrent and interlinked cycles of inquiry. The three cycles include: (1) the relevance cycle, (2) the rigor cycle, and (3) the design cycle (Hevner, 2007). The relevance cycle links the model to its contextual environment. The rigor cycle ties design activities to scholarly literature and established theory. The design cycle involves iterations of design, development, testing and evaluating the model. This successive process is also commonly referred as prototyping.

The entire research project is organised into eight stages. Stage 1 includes background research, question formation and project design. Stage 2 through Stage 6 covers the model design cycle and the prototyping processes of TI measurement technique. Stages 7 and 8 deliver the final quantitative and qualitative contributions of the dissertation. Findings and evidence presented contributes to a comprehensive productivity assessment of Australian universities. A visualisation of the research framework is given in Figure 1. The figure further indicates corresponding chapters of the dissertation that cover each research stage.

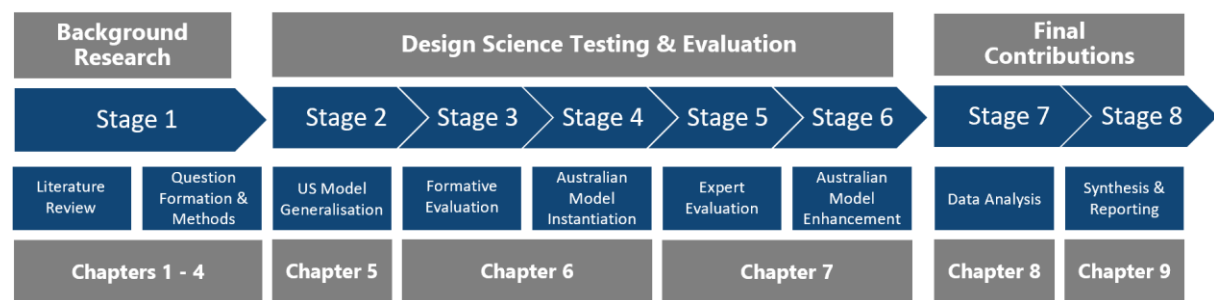


Figure 1: Research framework

### 1.4.3 Chapter Structure

The dissertation is organised into three parts. Part I covers the first stage of research, which incorporates the literature review, research question formation, and selection of methods. Part II encompasses the design science research cycle. It documents all model design, testing and evaluation procedures and their findings. Part III covers the final stages of research and contributions to knowledge and theory. It presents all findings and addresses the research questions and lists implications. Beginning each chapter, quotations are provided from individuals who participated in the interview phase of this research. Participants and their organisations have been anonymised, but further information about participants is provided in Table 27 in Section 7.2.2. The participant quotes were selected as indicative of the subject matter of each chapter. They do not appear elsewhere in the body of the dissertation. An abstract and key words are also provided for each chapter.

Part I contains Chapters 2, 3 and 4. Chapter 2 is focused on contextual issues for productivity measurement and performance assessment in higher education. It discusses global higher education trends and challenges with examples from Australian higher education. It finishes with multiple rationales for why improving university productivity measurement is important. Chapter 3 reviews the literature on productivity measurement and its application in the higher education sector. It introduces a taxonomy for identifying types of models for productivity measurement and uses the taxonomy to identify gaps in knowledge and research. It finishes by stating this study's guiding research questions. Chapter 4 is a shorter chapter that describes the research methodology and design chosen to address the research questions. The chapter explains how design-science research serves as this study's methodology and explains all methods used in this project.

Part II contains Chapters 5, 6 and 7. Chapter 5 is another relatively short chapter that covers research Stage 2. The chapter presents the scientific and scholarly backings for the first prototype measurement model in this research. It is based on a model first developed by the US National Research Council (NRC) for the United States higher education context (Sullivan et al., 2012). The chapter provides a rationale for model specifications and adaptations. Chapter 6 Explains how a collaboration with the Asia Productivity Organisation (APO) provides access to higher education leaders and practitioners from Cambodia, Fiji, India, Indonesia, Malaysia, Pakistan, the Philippines, Sri Lanka, and Thailand. These higher education practitioners use the prototype model to generate insights on university productivity. Learnings from the international applications of the model then inform model re-design for an Australian-specific instantiation of the model. Chapter 7 is structured similarly to the previous chapter. It provides all methods and rationales for testing and evaluating the prototype model. The methods, however, are distinct. This chapter provides results from interviews with 20 Australian higher education experts and stakeholders and how these individuals assessed the model to inform a full evaluation of the model. The chapter concludes with full details and explanations of the final model in the research.

Part III contains Chapters 8 and 9, which cover research Stages 7 and 8, respectively. Chapter 8 reveals numerous productivity trends evident within the Australian higher education system. It makes use of all techniques built into the final version of the model. The chapter unpacks and characterises many complex trends evident across institutions. Chapter 9 synthesises results from the quantitative analysis with qualitative evidence from the interviews. A synthesis of all results produces findings that contribute to knowledge, theory and practice. The research questions are addressed, implications are listed, and limitations are discussed. The chapter finishes with a discussion of conclusions and consequences of the research.

## **1.5 Summary**

### **1.5.1 Focus and Contributions**

The chapter began by explaining that the research in this dissertation seeks to extract as much value from quantitative data on universities as possible, but no more. It explained the benefits

of enhancing data-driven understandings of institutional performance and how those benefits outweigh concerns cited in the literature about the misuse of metrics. This chapter explained how the research is dedicated to finding a solution for generating meaningful information about universities that is guided by theory and suited to context. It seeks to do this using a design science methodology to design and test a new productivity measurement model using a calculation technique, called the Törnqvist index. The chapter further listed the key contributions of this dissertation, which include a new framework of paradigms for understanding how different stakeholders conceptualise and think about higher education productivity in context. Contributions also include technical guidelines, in the form of design principles, for improving university productivity measurement practice. Finally, key trends are measured in the Australian higher education system, and measurement results inform a performance assessment of the system. Assessment findings have implications for policy, practice and provision of higher education.

### **1.5.2 Considerations and Limitations Arising**

This chapter acknowledged that the scope of research does not include quality metrics, measures for the service and engagement function of universities, or measures for interdepartmental dynamics. While the research boundaries may preclude the ability to generate insight on certain key phenomena, the refined scope allows for more reliable findings for the phenomena it does target for measurement and assessment. The research further does not seek to capture the dynamics of non-university tertiary education institutions. The reader should also not conflate the terms productivity improvement and productivity growth with the term ‘innovation’. Innovation is a broader concept than productivity, and while a measured change in productivity may reflect some process innovation, not all innovations will have direct implications for productivity. This research is intentionally positioned to offer an alternative way of measuring institutional productivity in higher education, and it seeks not to add incrementally to the bulk of prior work on university efficiency measurement. Rather, it draws from a recent seminal work from the NRC to promote a directional shift in university productivity measurement practices.

PART I:

LITERATURE REVIEW AND RESEARCH DESIGN

## CHAPTER 2: THE CONTEXT OF HIGHER EDUCATION PRODUCTIVITY

*Who are the stakeholders, and what do they value? Parents want this.  
Students want that. Administrators want something else.*

–Study Participant TE.3  
University Professor and Economist

*We have rules that used to work. Now you get exceptions to the rules.  
And all the exceptions to the rule get added on until it falls over. You  
see examples of that everywhere in our sector.*

–Study Participant KS.9  
Academic Program director

### **Abstract:**

This chapter formulates an argument for why researching university productivity is an important scholarly and practical endeavour. The chapter argues that, while quantitative measures cannot capture all important aspects of higher education performance, currently available data represent prime opportunity. It is a missed opportunity from a scientific point of view, as well as from a practical standpoint with respect to the efforts and resources expended to collect, store, and report increasingly large amounts of information on public institutions. The argument is shaped and synthesised from the results of a literature review on the topic.

### **Key Words:**

Stakeholders, interests, trends, opportunities

## 2.1 Introduction

This chapter discusses the potential scope for insight of productivity measurement and assessment in higher education. It acknowledges how productivity analysis in higher education has been limited, but a review of the literature shows how further research on the topic could enhance policy and decision making in the industry, as well as advance scholarly knowledge. Productivity represents a ratio of output to input, but it is a deceptively simple concept. Productivity analysis is a complex endeavour in every industry (Tangen, 2005). The idea owes much to the fields of engineering and economics (Bairam, 1994; Sumanth, 1994; Gordon & Vaughan, 2011). Productivity measurement poses unique challenges in higher education due to the nature of university inputs and outputs, but the challenges can and should be addressed (Massy & Archer, 2018). Productivity assessment can help to track and understand the dynamics of a changing and high stakes industry, and measurement helps to form standards for improvement (Stufflebeam, 2003; Sullivan et al., 2012).

Challenges for accurately portraying university productivity stem from the availability of data for making calculations and from difficulties in representing the nature of the work being done within institutions (Singh, Motwani, & Kumar, 2000). In higher education, there are few conventions or standards for determining what data to use in a productivity measurement and how to treat the data for calculation. In the private sector, financial figures may be enough. For-profit firms strive to increase or preserve profits. Dollar figures thus speak to core objectives. Financial metrics for higher education and for broader public services, however, tell only part of the story of institutional performance (Linna, Pekkola, Ukko, & Melkas, 2010). Financial indicators do not speak to the central purpose of public educational institutions. Meeting financial objectives is critical for public institutions, but margin is secondary to mission in public universities.

Shortcomings associated with productivity measurement have fuelled criticisms, but this chapter demonstrates that measurement creates opportunities for insight that outweigh risks associated with data limitations. The insufficiency of financial data to represent university performance—paired with difficulties of capturing quality—has provided a basis for much criticism of productivity measurement in higher education. Arguments revolve around the misuse of metrics for performance management. Tendencies of metrics to distract and detract from important aspects of academic work. Finally, measurement is purported as a mechanism to shift agency and power to administrators and politicians. Political and ideological concerns cannot be separated from the provision of higher education, but the primary framing of productivity measurement in this chapter is one of empirical science and institutional research—not a framing of power, control, and trust. Measurement in higher education is difficult, but refined uses of measurement and new technological capabilities are transforming the field (Dooren & Van de Walle, 2008). Massy and Archer (2018) explain that pitting productivity assessment against the quality of academic work represents outdated thinking. When the purpose and discussion about productivity measurement revolves around

understanding and improving the experiences of students and enriching the environment for research—rather than around power, finances and austerity—meaningful measurement and assessment can be conducted without the threat of undermining quality and institutional mission.

The chapter is organised as follows to address the issues raised above. The chapter first defines and parameterises the topic of productivity in higher education. It then explains why studying university productivity is important both for practical purposes and for advancing knowledge about higher education performance assessment. The chapter then explains why productivity measurement has typically been a challenge in the higher education sector. It specifically discusses challenges with measuring and representing quality. The chapter then concludes by expanding upon rationales for conducting more research on the topic and suggesting how further research could target specific aspects of productivity measurement and assessment.

## **2.2 What is Higher Education Productivity?**

### **2.2.1 No Single Agreed Definition of Productivity**

A principle reason why more research on this area is needed is that no universal definition or common understanding of productivity in higher education exists. This chapter endeavours to generate a common, broad understanding of the term. It sets boundaries, and it clarifies what does not fall under the umbrella topic. The chapter intends to reduce ambiguity and to lay groundwork for framing a precise object of analysis for the remainder of the study.

Prior attempts to establish broad working definitions of the concept within higher education (Sullivan et al., 2012) have been met with pushback and controversy (Drengenberg & Bain, 2017). Tangen (2005) and Forrester (1993) explain that problems arise because the term, productivity, is both used widely in common language and frequently misunderstood. The result is that productivity is often dismissed as a viable phenomenon to analyse or improve. Productivity lends itself to measurement, but when measurement specifications are established, there can be winners, losers and unintended consequences. Misunderstandings often arise among stakeholders between concepts and the mathematical definitions of those concepts (Tangen, 2005). This chapter sets out to frame productivity in a way that reduces the scope for colloquial misunderstandings to confound research intentions. Two broad framings are provided in this section, a functional framing and a systematic framing.

### **2.2.2 Functional Framing**

The functional framing of higher education productivity addresses what core higher education activities or functions could be assessed in terms of their productivity. What are the key academic processes that take inputs and produce outputs? Higher education is multi-functional and multi-faceted. The primary functions, however, may be understood in terms of

collective framings of the purpose of higher education. The key purposes are commonly cited as generating new knowledge via research and disseminating knowledge via education and training (UNESCO, 1998). Scott (2006) also recognises teaching and research as the prime mission components of modern universities. Broader altruistic, cultural and societal functions are also recognised as essential parts of higher education's mission and are often categorised under a third service/engagement mission component. The service function of universities is one that brings benefits to external entities, to academic disciplines, or to the institution and its departments (Miller, 2007). Coates (2017) describes the service/engagement function as a broader 'propagation' of knowledge with outcomes and results not as immediately visible as those of teaching and research. The research in this dissertation focuses on education and research productivity. The exclusion of service and engagement is discussed in the next chapter, as well as chapters five and nine.

There exists a host of further ancillary and secondary functions that universities perform, which contribute to their overall productivity. Miller (2007) lists management, administration, and information services functions as key areas of work that support teaching, research and service/engagement. Financial management, course scheduling, compliance with laws and standards, human resources, vendor management, and information technology each fall under the list of secondary functions. These functions are not directly represented in a broad input-output framing of education and research productivity. Later chapters in the research, however, discuss how these secondary functions are essential and intertwined with research and education productivity analysis.

### **2.2.3 Systematic Framing**

A systematic framing of productivity views institutions from a process-oriented perspective, and positions productivity within a larger institutional performance ecosystem. Productivity refers to a relationship between inputs and outputs. Inputs and outputs exist within a broader performance environment that includes quality, customers, upstream and downstream systems (Sink & Tuttle, 1989; Slack, Chambers, & Johnston, 2010). In higher education, there exist innumerable ways to reduce universities into component parts. Clark (1983) stands as a seminal work in categorising and describing components of higher education systems. Borden and Bottrill, (1994) list three broad dimensions of higher education performance, resource utilisation, outcome achievement, and reputational strength. The remainder of this section relies on Miller's (2007) framework for examining HEIs. The framework rests on a foundation of five system elements that help distinguish aspects of productivity and broader higher education performance:

*There are five internal system elements: [1] leadership systems, [2] inputs, [3] key work processes, [4] outputs, and [5] outcomes. The leadership system provides overall direction and support to the organization. Inputs come from upstream systems in the environment and represent resources the system requires to carry out its work. Key work processes are tasks and activities that transform inputs into outputs. Outputs are products and services received or experienced by the system's customers. Outcomes are the intended results the organization seeks when customers experience or receive outputs (Miller, 2007, p38).*

Figure 2 contextualises the five elements by framing them within a linear, temporal illustration of university performance. The framework shows upstream systems as ‘feeding’ inputs into the key work processes of the institution. Those processes result in outputs, and finally, customers and external stakeholders are shown as interacting with institutional outcomes. The framework positions productivity within the broader institutional system. Inputs and outputs are shown to encompass key work processes, and the output to input ratio is a direct attempt to characterise those processes.

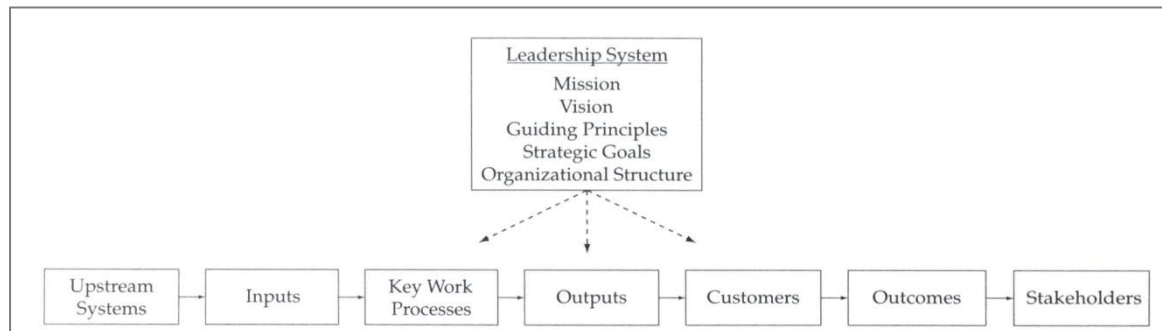


**Figure 2: Higher education performance system components (Miller, 2007, p. 38)**

It is helpful to illustrate additional performance areas that interact with productivity to better characterise the topic in context. Scerri and Agarwal (2014) note that linear representations of university processes are insufficient. They argue that academic work is not best represented in simplistic terms generally used for describing the production of tangible goods. An expanded framework is thus shown in Figure 3, which identifies seven interrelated and interacting aspects of university performance. The seven performance areas are described in Box 1.

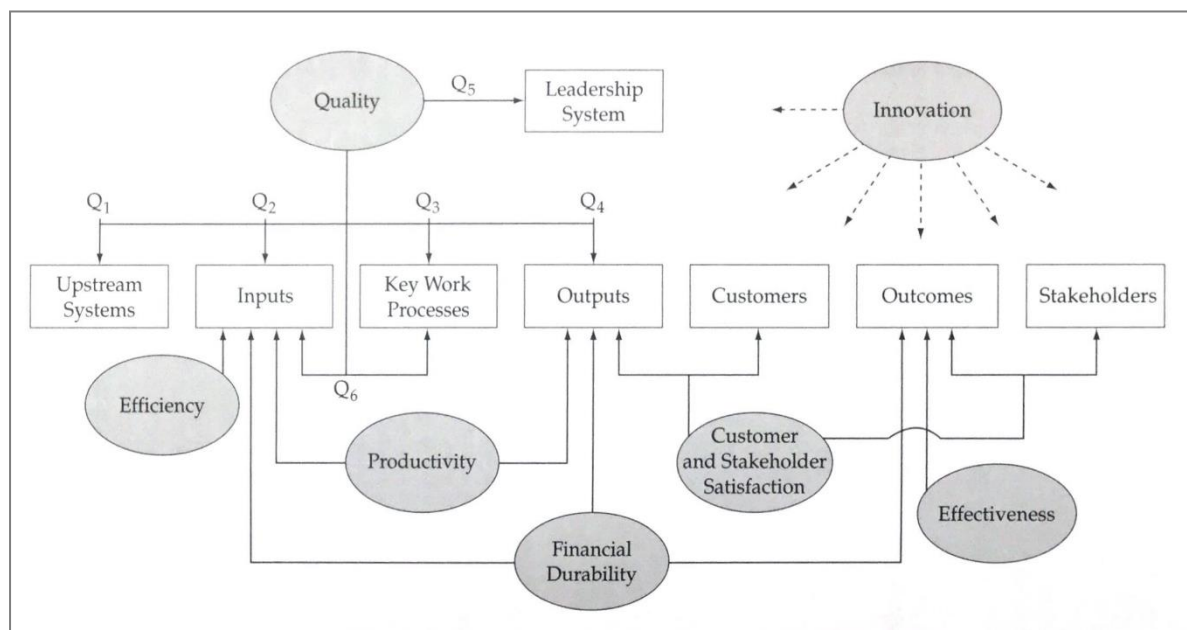


**Figure 3: Higher education institution performance system (Miller, 2007, p. 132)**

**Performance Areas:**

1. Effectiveness: a measure of the extent to which the unit achieves its intended outcomes
2. Productivity: a ratio of outputs created to inputs consumed
3. Quality: a complex area of performance measured in six dimensions:
  - Q1. Quality of upstream systems
  - Q2. Quality of inputs
  - Q3. Quality of key work processes
  - Q4. Quality of outputs
  - Q5. Quality of leadership systems
  - Q6. Quality of worklife
4. Customer and stakeholder satisfaction: a measure of the level of satisfaction of internal and external customers and stakeholders
5. Efficiency: a measure of resource utilization and the costs and benefits of quality management
6. Innovation: a measure of creative changes put into place to improve organizational performance
7. Financial durability: a measure of the organization's financial health

**Box 1: Higher education performance areas (Miller, 2007, p. 130)**

Productivity does not encompass all areas of higher education performance. It may be understood though as central to the performance system. Longer term outcomes of higher education and dimensions of quality are not considered as under the umbrella of productivity. Productivity in this dissertation is concerned with direct inputs and outputs of institutions, which encompass processes over which institutions have direct control. There do exist views that quality should be subsumed within the concept of productivity (Al-Darrab, 2000). Intuitively this makes sense. A measured, quantitative productivity improvement should not be considered a clear improvement if quality is diminished as a result. Practically, however, relationships between productivity and quality are difficult to demonstrate in a single measurement. Productivity and quality are thus frequently treated separately. Quality in higher education remains a prime productivity and performance issue, and the topic is discussed in more detail in Section 2.5.

## **2.3 Why is Productivity Important?**

### **2.3.1 Public Sector Accountability**

Measuring and understanding productivity in public higher education can address interests in accountability. Linna et al., (2010) list three connections between productivity and accountability in the public sector. First, public sector institutions are major employers in the workforce. Second, they provide many services in the economy that affect input prices and the quality of labour. Third, they are consumers of tax revenue. Recent decades have shown that concerns have grown about accountability for the provision of public higher education. Direct accountability to whom – i.e. a central government, fee paying students, or peer assessors – depends largely on context and country. However, a theme often associated with

accountability is on ‘value for money’ (L. C. J. Goedegebuure, Maassen, & Westerheijden, 1990). Questions about what precisely is ‘value’ and how much it should cost, however, are not easily answered. Further discussion about value as an output of higher education is provided in section 2.5.

In Australia, value for taxpayer money is a contentious issue in higher education, and ongoing work is needed to link costs to outcomes. Politicians make difficult decisions about prioritising public money, and voters have the right to ask whether resources are being used efficiently and effectively to generate results. In higher education, the cost side of the equation has received most of the attention in recent years. The 2003 Higher Education Support Act (HESA) outlined comprehensive funding reform, including provision for the country’s Commonwealth Grant Scheme (CGS). The CGS guaranteed public funding on a per student basis in public universities, rather than through teaching and learning block grants (DET, 2015). From 2012 to 2017 upper limits on the number of publicly funded domestic undergraduate students were removed with the enactment of the ‘demand driven system’. In 2018, however, funding payable through the CGS was capped at 2017 levels for each institution under arguments that the funding structure was unsustainable (DET, 2018c). New policy has yet to be enacted regarding the CGS, but a new funding model is scheduled to be finalised in 2019 (DET, 2018d). Funding arrangements and budgets are easily parsed, scrutinised and altered. Productivity analysis, however, is not zero-sum in relation to other budgetary items. It links inputs to outputs and provides characterisations of value creation. Productivity analysis can be used to justify resource allocation decisions on inherent merit, rather than by comparing the relative sizes or growth of budget items.

### **2.3.2 Emergence and Evolution of Trends across the Sector**

This section includes a significant review of scholarly literature, professional reports, and public data. It provides essential background information on broad trends in higher education. The section further presents a rationale for studying higher education productivity based on the need for institutions to understand these trends in relation to their own operating contexts. Aggregate indicators, such as productivity, always run risks of hiding details, but they promote understanding of broad phenomena. The sub-sections below illustrate how productivity analysis is relevant and appropriate across various issues concerning higher education’s evolving role in society. Each subsection addresses global trends and provides specific examples for Australian higher education.

#### **2.3.2|a. Scale of provision**

Higher education systems across the globe have been expanding to provide services, programs and facilities for increasing numbers of students. A prime use of productivity analysis is to understand the dynamics of growth and scaling (Besanko, Braeutigam, & Rockett, 2015). Literature and practice have not provided easy solutions for scaling higher education services and institutions. This sub-section on the scale of provision of higher education illustrates the growth that has been occurring internationally and in the Australian context.

Consider total student enrolment in higher education across the globe. The bulk of enrolments have tended to be students recently graduated from secondary school, aged 19-23 and are

often referred to as the traditional ‘college-age-cohort’. Diversity among enrolled students has been growing especially in rich countries, but students from the college-age-cohort represent the largest share (UNESCO, 2017). Schofer & Meyer (2005) note that in the year 1900, a total of 500,000 students were enrolled in higher education worldwide, representing only 1% of the college-age-cohort. Accelerated provision of higher education began in the United States and Europe post WWII, and continued throughout the second half of the 20<sup>th</sup> century (Trow, 2007). By the year 2000 there were approximately 100 million higher education students worldwide, representing 20% of the college-age-cohort (Schofer & Meyer, 2005). The total more than doubled by the year 2014, reaching 207 million and 34% of the college-age-cohort (UNESCO, 2017). By 2020, total global enrolment is expected to reach 205 million and as much as 375 million by 2030 (Calderon, 2018).

The story applies to Australia. In 1960 there were fewer than 80,000 students enrolled in Australia. By 2016 the number had increase to nearly 1.5 million. Most of the growth has been due to increased numbers of students progressing from secondary school. From 1989 to 2016 the percentage of 19-year-old students enrolled in higher education more than doubled from less than 20% to more than 40%. During the same period from 1989 to 2016, overseas student enrolment also grew in Australia from under 25,000 to nearly 400,000 (Norton & Cherastidham, 2018).

Post WWII has posed numerous challenges for higher education systems worldwide. Effects were being noticed as early as the 1970s. Trow (1973) explains that “the high growth rate places great strains on the existing structures of governance, of administration, and above all socialization” (p. 2). Not only have physical resources and administrations become more strained, but also, rapidly upscaling the size of the academic workforce has impacted academic work and culture. A prime concern is that the quality teaching and research is being adversely affected with focuses on efficiency and larger student to faculty ratios (Trow, 1999, 2007).

As higher education systems enrol larger proportions of the population, states face stark choices about how to fund institutions. In most countries across the globe, student contribution has increased. Globally, the proportion of public cost coverage of higher education has been decreasing relative to private cost coverage (Knight, 2003; Van Damme, 2015). Many countries in Europe have resisted imposing higher student fees. Trow (2007), however, warns against risks of underfunding higher education systems, stating that the productivity improvements required to educate more students and conduct more research at lower marginal costs are difficult to achieve and sustain.

Australia is a country that has increased requirements for student cost coverage over time. The Australian government introduced the Higher Education Contribution Scheme (HECS) in 1989, which allowed students to contribute to incurring a nominal debt without having to make any up-front payments. Adjustments were made to the HECS program throughout the 1990s, with substantial increases to student contribution imposed in 1997 (Jackson, 2001) and again in 2005 (Department of Education and Training (DET), 2015). The result is that, as student numbers have grown, individuals have assumed increased risk over the outcomes and promise of higher education.

### **2.3.2|b. Purpose of higher education**

As greater proportions of national populations hold stake in the outcomes of higher education, provision continues to evolve. Shifting purposes of higher education is another reason to be concerned about productivity. Olssen and Peters (2005) argue that since the 1980s, HEIs have been undergoing perpetual shifts in how they must justify their own institutional existence. In addition to issues mentioned above concerning performance at scale, productivity analysis is also commonly used to elucidate relationships between institutional performance and larger scale market changes or industry transformations (Olley & Pakes, 1992; Schiffbauer & Ospina, 2010).

Vingaard Johansen et al. (2017) explain how the role of higher education systems has evolved to drive the development of their country's knowledge economies and that universities have assumed the primary responsibility for doing so. Universities taking on this role represents a shift toward a more "instrumental" notion of higher education in support of pursuing national strategies to advance knowledge societies and knowledge economics (Kogan & Teichler, 2007). Santiago et al. (2008) show how nations spanning the globe increasingly recognize higher education's benefits as drivers of economic performance. Holland et al. (2013) further illustrate that higher degree attainment in a workforce is correlated with GDP growth.

The view of higher education as an economic instrument can also be seen at the level of individuals. In Australia and internationally, individuals are putting pressure on HEIs to elucidate the relationships between cost, quality, and expected career and lifetime benefits (Coates, 2017b; Van Damme, 2015). The idea that universities should be able to guarantee a measurable return on investment for students is becoming more widespread. A recent proposal for higher education policy reform in Australia stressed that higher education provision should be characterised by stronger and measurable connectivity to graduate employment outcomes (Australian Government, 2017).

### **2.3.2|c. Changes affecting the academic profession**

Through a productivity lens, academics may be considered as prime labour inputs who perform the key work processes of universities, which result in education and research outputs. Dynamics and variations within the academic profession serve as motivation to improve higher education productivity research. In their seminal work, *The Changing Academic Profession (CAP)*, Teichler, Arimoto, & Cummings (2013) assert that the academic profession has always been changing. They state that academic staff member's situations and activities are in constant flux. Bentley & Kyvik (2012) add that patterns of academic work and time spent on various activities differ significantly across countries, suggesting that conditions of academic work remain dependent on national or local higher education traditions. Traditionally, productivity studies have helped firms to understand evolving characteristics of inputs over time (Alcaide-López-de-Pablo, Dios-Palomares, & Prieto, 2014), as well as alternative configurations and usages of multiple inputs (Gürel & Kiliçaslan, 2016; B. Lin & Xie, 2014).

Despite variation across contexts, at least three prominent forces appear to be driving changes. Kogan and Teichler (2007) list the phenomena of 'relevance', 'internationalisation', and 'managerialism' as have been distinctly influential on the academic profession since the latter part of the 20<sup>th</sup> century. Relevance refers to pressure on academics to prioritize

developing demonstrable and transferable skills in students and research with more practical application. Internationalisation covers a range of topics and tends to ebb and flow in its centrality to global higher education issues. In this sense it refers to increased international mobility and flows of information, as well as the increased roles and functions of international knowledge networks. Higher education managerialism refers to university leaders' increased tendencies to borrow management and governance models from the private sector, which are more market-oriented (Kogan & Teichler, 2007).

In the Australian context each driver manifests in observable patterns. With respect to relevance, consider university reported research income. From 2011 to 2016, research income linked to industry partnerships grew most quickly across the sector. Income from traditional, competitive research grants that are publicly funded—and driven largely by researcher and academic interests—changed very little over the same period (DET, 2019b). Discussion about increasing numbers of overseas students enrolling in Australian universities illustrates internationalisation. Increased reliance on the overseas student fee revenue in Australia universities further emphasises the point (Croucher & Moore, 2018). Finally, growing numbers of contractual and casual academic staff across the sector indicate managerial approaches to human resources that are diverging from traditional practice. From 2007 to 2016 full-time and fixed-term academic staff grew by approximately 20%, while full-time-equivalent casual academic staff grew by approximately 60% (DET, 2019a).

### **2.3.2|d. Competition and league tables**

Institutional-level pressures related to competition and international rankings further support the improvement of higher education productivity analysis. Institutional rankings and competition present new realities for universities. Whether competition drives institutions to improve performance or whether competitive behaviour aligns with institutional goals is controversial and unclear (Marginson & van der Wende, 2007; Coates, 2016). Productivity analysis has often been used to understand effects of relationships and competition between institutions (Aiello & Ricotta, 2016; Kokko, 1996).

Hazelkorn (2015) laments how rankings do not capture the complexity of higher education. Further, criteria and methods for ranking are often not fully transparent. Yet, rankings still have great impact on perceptions and reputations of institutions. Rankings agencies create *de facto* incentives for particular institutional behaviours—whether those behaviours align with institutional goals or not (Marginson & van der Wende, 2007; Coates, 2016). International rankings agencies are also big businesses and may not be motivated to promote institutional behaviour that creates better outcomes for students and researchers. (van Vught & Ziegele, 2012) argue that improving transparency and increasing multi-dimensional representation within rankings is possible and describe their attempt at doing so with the creation of U-Multirank. It is also a free platform for users of the service and participating institutions, driven by epistemological interests, rather than corporate interests.

Australian universities have emphasised performance in international league tables, which has led to the country's quick rise in multiple international rankings. For the Times Higher Education rankings, the number of participating Australian universities rose from 7 to 35 between 2007 and 2018. For the QS World Rankings, the number of participating Australian universities rose from 25 to 37 between 2012 and 2018. For the ARWU rankings, the number of participating Australian universities rose from 14 to 23 between 2005 and 2018. But

Although Australian universities have risen in rankings, they tell a story of performance at indiscriminate cost, and performance on the cost side of Australian universities has been mixed (L. Goedegebuure & Marshman, 2017). The sustainability of financial trends is in question by government officials (DET, 2018). Productivity indicators fill both a practical and epistemological gap in understanding institutional performance dynamics by linking financial inputs and institutional outputs and outcomes most heavily weighted in global ranking schemes.

### ***2.3.2|e. Decoupling of Teaching and Research***

The decoupling of teaching and research in universities can be examined at an institutional level or at the level of academic staff. It refers generally to the separation of the core academic functions and specialisation in one or the other. It stands in contrast to traditional ideas of the university, under the Humboldt model, where teaching and research are considered as inseparable (Fallon, 1980). From a productivity lens, separating teaching from research, rather than thinking of teaching and research as being intertwined, represents an alternative functional and operational model for institutions. In the context of assessing the efficacy of operational shifts, productivity analyses have often been used to understand the impact of new functional dynamics, restructuring, and new internal strategies (Disney, Haskel, & Heden, 2003; Fujii et al., 2015).

Neumann (1996) describes the teaching-research nexus as a relic from the past in a new era of the ‘multiversity’. A dichotomy is growing between teaching and research. Segmentation of teaching institutions and research institutions, as well as teaching staff and research staff has become increasingly common (Altbach, Reisberg, & Rumbley, 2009; Nicholls, 2001). A prime concern is that the two functions have become imbalanced, and incentive structures favour performance in research over teaching (Coates, 2017b; Hazelkorn, 2015). A 2011 survey found that 80 per cent of Australian academics wanted to raise their research profile (Bexley, James, & Arkoudis, 2011). It is argued that these incentive structures have led to rational decisions by academics to prioritise research, and concerns have been raised about the ability of undergraduate teaching programs maintain quality in this environment (Fairweather, 1989; Harland & Wald, 2017).

In Australia traditional university values are upheld in national policy and funding frameworks. Under the Higher Education Standards Framework (Threshold Standards) 2015 minimum university standards for the conduct of research and for the provision of coursework are explicated (Australian Government, 2015). Yet in contemporary policy discussion, there is a push for more institutional diversity. The APC (2017) suggests that teaching and research in Australia are not best characterised as a nexus, but rather as two separate functions that compete for time and resources, as evidenced by common practices among institutions to subsidise research activities with revenues generated by education and training. It is argued, that new legislation could allow for a more diverse set of institutions holding university or equivalent status. A new breed of Australian university with allowance for a more targeted scope of provision on learning and teaching—and with access to public funds—could meet more diverse student needs.

If public HEIs in Australia continue to follow traditional institutional models amidst pressure to differentiate, however, decoupling of teaching and research continues to manifest within institutions. Consider trends in the contractual responsibilities of academic staff. From 2007

to 2016 the total number of full-time equivalent academic staff with both teaching and research responsibilities grew by only 7.5%. Over the same period the number of full-time equivalent teaching only staff grew by 77%, and research only staff grew by 32%.

## 2.4 Measurement Remains a Challenge

### 2.4.1 Opposing Views on Measurement

The study of higher education productivity has grown because of interests and pressures for institutions to demonstrate and improve performance, but many stakeholders remain critical about the value of measuring university productivity. Cave, Hanney, and Kogan, (1991) highlight that setting priorities and quantitative targets within higher education is difficult. They stress that an unrealistic consensus around priorities would have to be achieved for performance indicators to reach full functionality. The multi-component, multi-input, multi-output, and multi-purpose nature of higher education muddies the waters of analysis and interpretation of productivity measurement results. Societal, government, and stakeholder expectations, though, have shifted toward institutions finding measures, making claims, and providing evidence to justify their relevance and outcomes (Bok, 2005). And increasing technological capabilities and data availability since 1991 allow for increasingly precise measures.

Olssen's and Peters' (2005) and Woelert & Yates (2015), though, warn of risks associated with the notion of higher education as a mechanism to produce narrowly defined, discrete outcomes. They paint a picture of academic values being eroded and the quality of academic work being degraded because of commercial interests, market forces and performance measurement. Drengenberg & Bain (2017) add that contemporary understanding of productive processes and practices in higher education is not yet sufficient for productivity indicators to provide meaningful information and feedback. They argue that currently available data is unable to give a valid portrayal of productivity.

Alternatively, Brennan (2007) does not view pressures on institutions to demonstrate outcomes and performance an existential challenge to traditional values. The author argues via *reductio ad absurdum*, challenging stakeholders to imagine a world where, instead of relevance, society calls for higher education to be *irrelevant*. Massy (2016) also frames challenges to the status quo as neutral realities. New challenges simply create imperatives for continual improvement of higher education's core services and value propositions—and measurement is necessary for gauging progress. Further, there exists a growing literature on higher education productivity measurement and its improvement. Studies in this literature weigh the implications of alternative technical models (J. Johnes, 2006). Others have demonstrated sufficient confidence in quantitative measures to present comparative efficiencies between individual universities and system-wide productivity growth to a level of precision at the tenth of a percent (Moradi-Motlagh et al, 2016)

Reasonable individuals disagree about whether measuring university performance should be viewed in a pessimistic or optimistic light. Yet, even if quantitative performance portrayals

are limited, neglecting available and reliable data represents only a missed opportunity to learn from the efforts and resources spent to collect, store, and report information about universities. The information contained in legally required reports and publicly available data bases reflects what key stakeholders and the public have deemed important. Quantitative indicators do not represent all that is important, nor perhaps even a majority of what is important, but that is not the burden of such indicators. Development of productivity metrics should not strive to become more comprehensive in scope but instead strive to reduce misrepresentation or misinterpretation. That higher education contexts and challenges are entrenched in interests and values means that measurement can never be values free. But improving measurement with the development of transparent measures that are amenable to change as the provision of higher education changes can generate far more information than misinformation.

#### **2.4.2 Politics, Power and Trust**

Even if challenges of misrepresentation or misinterpretation of measures can be overcome, there still exist challenges of misuse. Dooren & Van de Walle (2008) note that any performance measurement scheme will have a political and power dimension. Once school of thought views performance measures through a strictly political lens. Performance measurement has been cited as a symptom of neo-liberal philosophies and New Public Management (NPM) (Feller, 2009; Kallio & Kallio, 2014; Olssen & Peters, 2005). The arguments associate performance measurement with ‘measurement regimes’ that are steeped in distrust between academics, administrators and the state. Performance measures are framed as tools to control the work of otherwise intrinsically motivated academic staff (G. Forrester, 2011). Measures are thought to represent central authorities’ assumptions and expectations that universities default to inefficient and ineffective states of operations (Woelert & Yates, 2015). Further, Olssen and Peters (2005) claim that NPM is fundamentally incompatible with traditional approaches to higher education provision. That is, higher education upholds notions of ‘public service ethic’ and ‘common good’, and NPM and performance measurement undermines those notions. Alach (2017) notes that, given the general condemning tone on higher education performance measurement in the literature, one should anticipate that any quantitative study of university performance would be met with little support by academic staff in terms of both implementation and benefit.

Yet political arguments in support of performance measurement continue to drive the status quo. Higher education performance measures have been—and continue to be—used for high stakes decision-making and to hold universities to account (Alexander, 2000). Barnett (1992) justifies the use of performance assessment, stating that “society is not prepared to accept that higher education is self-justifying and wishes to expose the activities of the secret garden” (p.16). Cave et al. (1991) argue even further to the contrary that performance measures can be used as mechanisms for *increasing* institutional freedom from central powers. With appropriate indicators, central authorities may be concerned with only end results, rather than meddling in the academic processes used to achieve results.

This dissertation recognises and addresses political issues, but the primary framing of productivity measurement is one of empirical science and institutional research—not a framing of power, control, and trust. In recognizing that higher education performance

indicators continue to be used for decision-making, the imperative for inquiry is to continually improve indicators for accuracy, reliability and relevance. This dissertation does not align with politically grounded arguments for or against measurement. It recognises that there can be winners and losers under measurement regimes. The objective is, thus, to improve measurement practice as a science by acknowledging diverse stakeholders interests in pursuit of improving universities for the benefit of the societies in which they are situated.

### **2.4.3 Overcoming Challenges**

Measurement in higher education is difficult, but refined uses of performance measurement and new technological capabilities are transforming the field (Dooren & Van de Walle, 2008). There remain perceptions that performance measurement is equivalent to ‘performance management’ or ‘management by numbers’ (Alach, 2017). Such views often pit the practice of improving measurement and productivity against the maintenance of academic values and quality. But this dissertation attempts to argue that the pursuit of both productivity and quality does not amount to a zero-sum-game. Massy and Archer (2018) assert that pitting productivity against quality represents outdated thinking. When the purpose of productivity measurement revolves around enquiry and improvement, rather than power and efficiency, meaningful measurement and improvement can be made without undermining quality. Issues of quality are addressed in more detail in the next section.

Overcoming challenges related to quality, however, is not straight forward. An appropriate starting point may be to acknowledge and remain aware of Goodhart’s Law, which states that once a measure becomes a target, it ceases to be a good measure (Elton, 2004). The logic revolves around the idea that to satisfy new performance targets an organisation or individual must generally diverting resources from other essential, unmeasured tasks to produce desired results in the measured dimension. The practice often leads to unintended, negative consequences. In a recent book, “The Tyranny of Metrics”, Muller (2018) laments how metrics have been used as behavioural incentives and performance targets before having been vetted as reliable indicators of real phenomena. In doing so, the author acknowledges that metrics can serve as powerful diagnostic tools when measurement objectives revolve around problem identification and investigation. Dooren & Van de Walle (2008) emphasise that—with growing data collection and availability—the perception and use of performance measures as diagnostic tools in organisations is increasingly common. Advances in computing power and tools for data interpretation and visualisation have allowed organisational measurement to provide meaningful and valid insight.

Overcoming measurement challenges related to power and trust is also complex. Such challenges often manifest in relationships between trust and transparency. Trust and transparency are often pursued in tandem under assumptions that trust is built through transparency. In the context of productivity measurement, transparency may refer to the release of new data not previously reported. In the case of Australian higher education, the DET (2017) states, “The higher education sector will be more accountable—and more open to scrutiny—for the manner in which it expends taxpayer funding. There needs to be greater transparency to show that the institutions are making the right decisions.” The statement outlines how expenditures on teaching and research are being investigated and that the government will work with universities to set terms for publication of findings. Such

information would inherently improve the reliability and the accuracy of productivity metrics for the research and education functions of universities. Botsman (2017), however, defines trust as confidence in the unknown, and the author laments that calling for the release of data is often an inherent instigator of problems concerning trust and power. The practice of performance measurement in public universities will always have to balance trust and transparency and weight the implications of mandating the release of data.

Cave et al. (1991) and the OECD define a performance indicator as “a numerical value used to measure something which is difficult to quantify” (p. 21). Any measurement exercise holds risks associated with rigor and transparency. Measurement results may always be subject to scrutiny or dismissal by stakeholders with opposing interests. Debate and research, however, should focus on how best to conduct performance measurement in context and how much weight should be given to measurement results in decision-making. Debate should not entertain claims about inherent merits of quantitative information and its suitability for communicating aspects of university performance. Quantitative measures by themselves cannot encompass everything that is important for all stakeholders—nor can qualitative indicators or any other single type of information. Parallel, supplementary, and multi-faceted analyses in higher education will always be needed to allow for qualitative and quantitative results to inform one another (Sullivan et al., 2012).

## **2.5 How to Address Quality**

### **2.5.1 Treatment of Quality**

For universities, quality is inherent to productivity (Linna et al., 2010; Sullivan et al., 2012). If a higher education system or institution is growing, then productivity improvement is contingent upon scaling quality practice. As discussed in section 2.2.3, sacrificing quality for efficiency does not merit a productivity improvement. Atkins and Herfel (2006) stress that when a productivity improvement agenda focuses only on inputs, quality and overall performance are at risk. Contentions further arise when conceptual understandings of quality and performance are not matched by measurement capabilities. If definitions, objectives and measurement are not well-aligned, the productivity targets may degrade quality (Ammons, 2004; Drengenberg & Bain, 2017). Even if imperfect, however, productivity measures retain inherent value. For Australian higher education, productivity measurement has the potential to both reflect and supplement the system’s established quality assurance and assessment systems.

When quality is independently and accurately assessed, productivity reflects quality. The work in this dissertation rests on Australia’s existing quality infrastructure and legal requirements regarding institutional standards that all universities must maintain. Quality standards are outlined and regulated by the Tertiary Education Quality Standards Agency (TEQSA) (Australian Government, 2011). TEQSA assesses and accredits institutions with a guiding objective to promote and enhance quality in higher education. This dissertation does not seek to critique or defend the work of TEQSA. It is assumed that the inputs, outputs and

processes of accredited Australian universities are adequately assessed and vetted for quality. Productivity measurement is thus conditioned on quality. Independent quality assessments underpin the validity of quantitative productivity representation (Sullivan et al., 2012).

That productivity indicators partially rely on quality assessments, however, gives rise to notable limitations of any independent productivity measurement exercise. Measurement can and should be improved to better reflect variable and changing quality dynamics. University input and output quantities are easily tracked on an annual basis, but comprehensive quality assessments cannot be performed so regularly. If annual fluctuations in productivity indicators are to reflect the effectiveness of processes and the quality of deliverables, some aspects of quality must be present in annual productivity measures. Options for doing so may be limited, but the challenge is not insurmountable.

### **2.5.1 Capturing and Tracking Quality**

Higher education quality is difficult to characterise. There exists no single, common or agreed understanding of the phenomenon. Elken & Stensaker (2018) review and describe the diverse body of work on higher education quality. Two conclusions emanate from their work. The first is that the lexicon is fragmented with few overarching concepts. The second is that the lack of common analytical tools to investigate quality further makes conclusions about the phenomenon difficult to make. Coates (2017) adds that quality in higher education is difficult to assess because of its nature as a ‘credence good’. Credence goods are goods or services where the seller or provider knows more about the quality than the consumer, and thus, quality is difficult to judge even after purchase or consumption (Dulleck & Kerschbamer, 2006).

The elusive nature of higher education quality is at the heart of why its incorporation into a productivity measure will always be difficult. Some unidimensional indicator must be selected, and limitations must be acknowledged. A singular, quantitative portrayal of quality cannot appease a diverse set of performance interests, but if refined, it could add scope and reliability to a numerical productivity indicator. The prime challenge for advancing productivity research in higher education is to pay some credence to quality while retaining reliability and objectivity, so that blanket assumptions about universal and constant quality do not have to be made about all institutions accredited by a single agency.

### **2.5.2 Quality, Productivity and Value**

The concept of value is itself complex but useful for the practice of higher education productivity measurement. Value is recognised as being linked to quality. It refers to direct benefits of a product or service, as experienced by customers or users (Tellis & Johnson, 2007). All else held equal, quality and value should be positively correlated, and quantitative proxies for value may be easier to identify than those for quality. Price is the quintessential example of a proxy for value in the private sector. Productivity analysis in the private sector has accordingly been refined to incorporate input and output quantities, along with price, in order to approximate the total value add of productive processes (Bairam, 1994). The concept of value is also helpful in the study of higher education productivity. Massy (2016) observes that, rather than profit, universities exist to produce value. They create value for

students, for academic disciplines, for communities, and for other stakeholders and partners. If output quantity is increased without demonstrably increasing value to students, researchers and other stakeholders, then output growth cannot be considered a productivity gain.

In higher education the best currency to express value is not dollars. Although purchases and sales occur in higher education, key higher education outputs are difficult if not impossible to price. Knowledge, service and engagement cannot be valued reliably using dollars. Yet, the study of productivity operates under the principle that the purpose of productive processes are to generate outputs with more value than raw inputs, regardless of how difficult it is to characterise value. Massy & Archer (2018) and Moore (2018) each describe how value in higher education can be expressed without heavy reliance on financial data. The authors focus largely on student success factors. As with quality, no single indicator could comprehensively encapsulate value in every context. But as the practice of productivity measurement has evolved to incorporate value, determining higher-education-appropriate indicators could enhance university productivity measurement. Section 3.3 in the next chapter adds more nuance to the discussion, and Section 5.4 in Chapter 5 addresses value directly by delineating how the concept is represented in higher education productivity measures in this research.

Assessing and indicating value and quality will always be limited. As stated above, comprehensive investigation of performance requires parallel studies of productivity, quality and other matters of interest in context (Miller, 2007; Sullivan et al., 2012). Productivity indicators can represent a discrete set of pre-specified performance dimensions, but no assessment of productivity can stand alone as a definitive indicator of overall performance. The prime shortfall of higher education productivity analysis is that current methods and results have not yet instilled a critical mass of confidence in stakeholders. Productivity assessment must be further developed to stand on its own without having to rely on parallel quality assessments for baseline validity. The next section further explicates how measurement can and should be improved.

## **2.6 Why and How to Improve Measurement**

### **2.6.1 Reasons to Improve**

A prime conclusion of the literature review in this chapter is that the opportunities and benefits of researching and enhancing data-driven understandings of institutions outweigh concerns about the misuse of measurement. Quantitative observations can both reflect and inform policy, practice and performance. The sections above delineated the ‘what’, ‘when’, ‘who’ and ‘where’ of productivity measurement in higher education. We now cover the ‘why’ and ‘how’ for enhancing its continued development. This section concludes the argument for why improving approaches to measuring higher education productivity are important both practically and theoretically. The reasons why also present broad avenues for how to do so, which is discussed in more depth in the next chapter. In summary, the contexts above illustrate that productivity indicators need to be improved for validity, reliability and

relevance in context. Such improvements could allow for increased utility and insight into university performance—as productivity measures have been able to do in other industries.

Better productivity indicators are necessary, not just for understanding but also for improving performance and productivity. Productivity assessments can speak to input-centric interests, such as funding allocation and finances, and output-centric interests, such as the quality and volume of the products of academic work, providing a wholistic picture of institutional performance. Improving productivity measurement means becoming more confident in identifying and articulating what is important to improve. Measuring what is valuable, rather than valuing what is measurable, should be the guiding principle in learning about how to achieve better outcomes with less consumption. The basic desires to innovate, improve, and increase productivity are shared across sectors and industries. Misuse and misinterpretation of measured productivity, however, poses the challenges for measurement to be used in earnest at the institutional level. Productivity improvement targets are likely not the best way to define performance policy, but improved measurement can reveal novel insight about institutional and system-wide performance trends. Better and more nuanced understandings of issues and trends beget better and more nuanced definitions of problems. And how we define a problem leads to how we define a solution.

### **2.6.2 Measured Results Should Align with Observable Behaviour**

Higher education productivity indicators should instil confidence that measured performance is the result of authentic institutional behaviour. Measurement approaches need improvement to ensure that estimates are not the result of idiosyncrasies associated with specific calculation techniques or the result of intentional manipulation of source data. The complexity of productivity and the aggregate indicators that represent it need to be parsed and evaluated from multiple angles to expose where risks lie in misrepresenting institutional productivity portrayals.

With notable scepticism surrounding the use of performance measures for higher education in general, productivity metrics need intentional investigation and development for their utility as a practical, institutional diagnostic tools. There is a gap that needs filling with respect to the diagnostic utility of performance measures in applied research (Sullivan et al., 2012).

Risks also need to be exposed for how productivity metrics could be manipulated. The literature review for this chapter did not uncover a large body of work on the ‘gaming’ or manipulation of institutional metrics in higher education, but many concerns and commentaries exist on the topic. Gruber (2014) highlights the vulnerability of research metrics to be used as tools for personal and institutional gain without reflecting true advances in research. J. Lin (2012) explores explicit anti-gaming mechanisms for research metrics, increasing resistant manipulation. Transparency in the construction of a metric is chief among potential options. Research indicators serve as essential data elements of many productivity input-output frameworks, and analogous concerns remain for other data elements included in specific input-output frameworks. Further limitations should be explored with respect to aggregating such data elements.

### **2.6.3 Productivity is a Lead Indicator for Longer-term Achievement**

Productivity is an indicator of longer-term success, but how to interpret relationships between inputs, outputs, and outcomes is not straight forward. Outcome achievement has become an increasingly important strategic objective of institutions. This section explains how improved productivity measures could indicate how well positioned an institution is to fulfil its longer-term promises. Recalling Miller's (2007) performance assessment framework from Section 2.2, outcomes are observable only after customers and beneficiaries interact with university outputs. Outcome achievement depends upon how students, communities and users of research benefit from outputs in broader societal and academic environments. Although such interactions are largely outside the control of institutions, success with outcomes has come to represent accountability to stakeholders and the public. By characterising within-institution functional dynamics, productivity trends allow decision makers to balance short-term dynamics with long-term goals, as well as analyse trade-offs associated with different performance dimensions.

Because productivity encompasses the range of activities that take place within an institution, productivity measures have been recognised as key diagnostics for gauging success across performance dimensions. Sink & Tuttle (1989) explain that productivity metrics can signpost efficiency, quality and effectiveness associated with outcomes. The implication is not that higher productivity indicates a higher likelihood of outcome achievement. Rather, productivity dynamics can signal trade-offs between inputs, outputs, notions of quality and strategic objectives. The extent to which this is possible depends upon what data and information are considered for productivity estimation. Data elements must be chosen and vetted as reliable proxies for key intermediaries that serve as prerequisites for generating intended outcomes. Enhancing the understanding of strengths and limitations of different combinations of data elements for doing is where further research has a role to play.

### **2.6.4 Analytical Framings of Productivity Need Refining**

There exist few conventions or parameters in higher education around choosing data for productivity analysis, interpreting data in terms of inputs and outputs, and choosing techniques to make calculations. First, a body of literature describes dozens of potential institutional indicators that could be used for higher education performance assessment (Alexander, 2000; Cave et al., 1991; Coates, 2016; De Witte & Sneyers, 2017; Miller, 2007; Paige, 2005). Choices, though, often rest on what data are accessible as much as on what analytical objectives need addressing. Some studies in the literature compensate by taking a comprehensive approach with available data. Wolszczak-derlacz & Parteka (2013), for example, synthesise measurement results from four different input-output frameworks to generate insight beyond what a single framework could provide. The situation still speaks, however, to the current *ad hoc* nature of university productivity measurement.

Data interpretation and specification as inputs or outputs within analytical frameworks needs attention. Practice is not consistent, and implications of different approaches in different contexts is difficult to surmise. For example, Yaohua, Muyu, Weihu, & Xianyu (2018) designate students as outputs in their productivity analysis, but Wolszczak-derlacz & Parteka (2013) specify students as inputs. Students may intuitively seem like inputs because they are admitted to an institution as participants of planned educational processes. Students may,

however, be considered as outputs if their numbers serve as a proxy for the scale of educational services delivered by an institution (Sullivan et al., 2012). Chapter 3 discusses higher education inputs and outputs in further detail and provides specific opportunities for enhancing the state of the art of productivity measurement.

Few conventions exist around quantitative techniques to generate productivity estimates. Concerns about what types of assessments are most appropriate in what contexts were raised by (Cave, Hanney, and Kogan; 1991), and answers up to now are still unclear. Singh et al. (2000) summarize three umbrella methods to estimate productivity across industries: linear programming, econometric models, and index measures. Linear programming is most common (De Witte & Sneyers, 2017), but all methods have unique constraints and opportunities for insight. The three techniques are discussed in more detail in the next chapter. Research and evidence-based decisions on appropriate analytical frames and techniques for productivity analysis are increasingly necessary. With the volume of data that higher education institutions collect, store and report constantly increasing, and with increasing computing power available to examine it (Dooren & Van de Walle, 2008; Hazelkorn, 2015), the diversity of frameworks and methods will only increase and thus need ever stronger justifications and rationales for use.

### **2.6.5 Alternative Metrics are Needed at the Institution-Level**

Productivity metrics provide an alternative to global rankings for characterising institutions and their performance. They have a different purpose, a different intended audience and are associated with a different set of incentives. First, the sector's performance benchmarking purposes and approaches need attention. Global rankings create a standard of comparison for institutions, but individual rankings indicators only vaguely represent authentic aspects and concepts of performance. Rankings are relative, moving targets with capped values and with little inherent meaning. Gaynor (1998) explains that standards are points of reference against which problems are defined and solutions are measured. Rankings do not fit this description. The author continues that the purpose of measuring institutions is to define gaps between current circumstances and desired circumstances. Productivity metrics speak to actual circumstances, operations and operating contexts. They speak to tangible aspects of performance and when parsed can inform varied performance issues and interests. Yet, international rankings currently provide stronger incentives to adjust performance and refine operations.

The rise of international rankings drive institutional behaviour because they have come to represent broad assumptions for what should be counted in terms of university performance. Dearden et al. (2008) show that contemporary university behaviour is best explained through the assumption that institutions are functioning to maximise rankings—rather to maximise contributions to their missions or to the public good. Even if these findings are not representative of all institutions, rankings-influenced behaviour is broadly observable (Hazelkorn, 2015). Increasing focus on international rankings—which have limited instructive value for making within-institution decisions and interventions—signals the need for alternative data representations more aligned with processes and behaviour to balance performance incentives (Coates, 2016).

The intended audience of productivity assessments is institutional leaders and decision-makers. The prime intended users of rankings are potential students deciding where to apply and enrol. Yet rankings have crept into the decision-making criteria of institutional leaders. Productivity indicators, however, could provide more valuable information and allow for better explanation of the nature and shape of institutional trends. Massy (2016) argues that more rigorous business analysis is needed by leaders in higher education. The author concedes, though, that specific methods should not emulate those of private sector firms. Public and not-for-profit universities have different business models. Rather than seeking to maximise profit, they should function to maximize mission over margin. Specialised and more nuanced analytical approaches are needed for leaders and practitioners to address higher-education-specific business questions. Astin (2012) further notes that institutional performance assessment and the use of comprehensive performance indicators have traditionally been used for accreditation activities. Practical use of measurement has been neglected within institutions. Determining and characterising new measurement and assessment approaches is the role of further research on productivity.

## **2.7 Summary**

### **2.7.1 Focus and Contributions**

The chapter provided a foundation for the remainder of this dissertation in three key ways. First, it framed and parameterised the concept of productivity in higher education. Second, it provided rationales for conducting further research on productivity and its measurement. Third, it discussed controversies, compromises and practical options for implementing further research. Higher education productivity was framed as a dimension of institution-level performance for universities. The chapter then parameterised the topic with respect to what university functions could be measured in terms of their productivity. Designated functions were limited to education and research, though the chapter did acknowledge other university functions, such as service and engagement. Productivity was also positioned as only one measurable dimension of university performance. Productivity was acknowledged as a central dimension of performance, as it seeks to encompass and represent the efficiency and effectiveness of institutional key work processes.

Rationales for engaging in further study of higher education productivity were provided. They include (A) the range of institutional and contextual issues that productivity can inform and (B) direct examples of current shortcomings of productivity measurement. Institutional and contextual issues where productivity analysis can provide insight include, institutional growth and system expansion, societies' evolving expectations for universities, the changing nature of academic work and university functions, as well as institutional competition and ranking. Current shortcomings to address with further research involve: linkages between quantitative indicators and observable, authentic behaviour, the ability to assess business processes and gauge progress toward long-term objectives, as well as contextually appropriate and consistent analytical framings productivity.

### **2.7.2 Considerations and Limitations Arising**

This chapter acknowledged legitimate reservations about measuring the performance of universities, including the potential for misuse of metrics in performance management. This dissertation recognises and addresses political issues, but the primary framing of productivity measurement is one of empirical science and institutional research—not a framing of power, control, and trust. Nevertheless, productivity assessment will always have limitations because of assumptions that must be made about data and measurement to generate estimates. A compromise in current measurement methods regards how productivity analysis incorporates quality. Quality must be assumed as inherent to productivity because efficiency at the expense of quality cannot be considered a productivity improvement. The implication is that, to some extent, productivity analyses must always be vetted and verified with parallel quality assessments. However, in moving forward with research, there do exist tested methods from other industries for how productivity measures can capture aspects of quality and value of outputs. Accounting for value and quality, along with other options mentioned above, illustrate how this research plans to contribute to the forefront of knowledge on higher education productivity measurement and assessment.

## CHAPTER 3: CONCEPTS AND TECHNIQUE OF PRODUCTIVITY MEASUREMENT

*The problem is we don't really measure what we want.*

–Study Participant TE.2  
Education Consultant

### **Abstract:**

The prime contributions of this chapter are (1) to highlight gaps and opportunities for improving higher education productivity measurement and (2) to discuss the research questions prior to formulating the methods used to address them. Insights in this chapter are generated from a literature review on the topic. It first summarises key concepts associated with the frame of production theory as it applies to higher education. A core finding of the literature review in this chapter may best be summarised by De Witte & Sneyers (2017): In higher education, “the efficiency literature pays currently a significant effort to optimize its own models. In doing so, it focusses on minor methodological details and neglects the current important issues in related fields” (p. 356). A prime purpose of this chapter is thus to provide clear directions for further research that are more sensitive to broader measurement issues and local Australian contexts. A prime direction identified for technical research is the development of indexing methods for measurement, as opposed to methods based on linear programming and data envelopment analysis.

### **Key Words:**

productivity change, production frontier, Törnqvist index, data envelopment analysis

## **3.1 Introduction**

A range of approaches exists for measuring and analysing institutional productivity. Common techniques, however, are too limited for the context of higher education. Different approaches are associated with unique strengths and limitations. This chapter explains how common productivity measurement practices in higher education have been borrowed from other disciplines and industries without full consideration of their implications in the new context. Borrowing techniques has practical benefits for the feasibility of conducting measurement and for rendering analysis. The scope for insight of borrowed techniques, however, may be misplaced. This chapter thus explores how alternatives to the norm may provide better contextual suitability and may increase the accuracy of estimates, even if they sacrifice statistical precision.

This chapter argues that while common measurement procedures provide benefits in terms of their readiness to handle complex datasets, they often produce results that are often too positivist and preclude opportunities for alternative interpretations of data. The subjectivity of institutional input and output data in higher education requires a softer, albeit rigorous, approach to measurement. The subjectivity inherent to higher education data cannot be eliminated, but students, government officials, university leaders and taxpayers have the right to ask whether the outcomes of higher education justify the costs of running the system. Increasing the value of productivity assessment and its potential to inform policy and practice, however, likely hinges better foundations for its measurement.

The chapter is organised as follows. The chapter begins with a summary of the microeconomic theory of production. The chapter then covers underlying assumptions associated with using Production Theory as an analytical frame for research. It then provides a history of how production theory and related economic concepts have been applied to research in higher education. A more nuanced discussion about productivity measurement models for higher education then follows. The chapter then explicates and categorises areas for further research on the topic and how it may be conducted. Finally, the chapter concludes with a detailed rationale for the three research questions presented in the introductory chapter.

## **3.2 Production Theory**

### **3.2.1 First Principles**

Production theory provides an analytical framework for exploring the concept of productivity—how scarce resources are used in productive processes to generate outputs. The

theory rests upon a perspective that organizations are socio-technical systems that transform inputs into outputs by means of their technological capabilities. Technology in this sense refers to any human, physical or information system performing key work processes. The intentionally broad interpretation of an organization's production technology allows for a wide range of applications across industries and organizational types, including higher education and universities. Representation of production technology helps to clarify relationships between inputs and outputs and may characterise implicit or explicit rules that explain possible input and output combinations and behaviours between them. These rules are embodied in the production function (Antle & Capalbo, 1988; Gordon & Vaughan, 2011; Hill, 2017). Production functions are explained further in the next section.

It is helpful to position production theory within a larger realm of organisational analysis. Production theory directs attention toward factors endogenous to an organisation. It attempts to characterise essential internal value-adding functions and mechanisms inherent to a firm or institution's scope of work. In some respects, production theory may stand in contrast to institutional theory. Institutional theory directs attention toward forces exogenous to the organisation. Institutional theory asks how organisational choices arise and how decisions are made within broader a institutional environment and context (Hoffman, 1999; W. R. Scott, 2008). Exogenous factors must be acknowledged in the study of productivity, but explicit analysis draws boundaries around the direct inputs and outputs to a firm's or institution's production technology (Färe, 1988). The term, technology in this sense, refers to the human and technological systems functioning to transform and add value to inputs.

Even with a bounded scope of analysis, many complexities arise. To establish a shared vocabulary, differences between production and productivity should first be unpacked. Used in isolation, the terms represent separate ideas and are easily confused. Production usually refers to the amount of output produced. Productivity, however, is relative with respect to input (Tangen, 2005). Box 2 and Box 3 illustrate how different input-output relationships that can result in result in improvement or decline in productivity. Further, in each example, attributing the behaviour of the inputs to the outputs, and vice versa, gives five instances of *decreased* productivity. The prime implication is that increases or decreases in production may or may not indicate corresponding increases or decreases in productivity.

1. Effective scaling (scaling up benefits)

$$\frac{\text{Output (increases faster)}}{\text{Input (increases)}}$$

2. Improving effectiveness.

$$\frac{\text{Output (increases)}}{\text{Input (constant)}}$$

3. Full innovation.

|                          |                                                                     |
|--------------------------|---------------------------------------------------------------------|
|                          | $\frac{\text{Output (increases)}}{\text{Input (decreases)}}$        |
| 4. Improving efficiency. |                                                                     |
|                          | $\frac{\text{Output (constant)}}{\text{Input (decreases)}}$         |
| 5. Efficient downsizing. |                                                                     |
|                          | $\frac{\text{Output (decreases)}}{\text{Input (decreases faster)}}$ |

**Box 2: Five relationships for increased productivity (Mistereck, Dooley, & Anderson, 1992).**

|                                             |                                                                     |
|---------------------------------------------|---------------------------------------------------------------------|
| 1. Inefficient scaling (net expense growth) |                                                                     |
|                                             | $\frac{\text{Output (increases)}}{\text{Input (increases faster)}}$ |
| 2. Declining effectiveness.                 |                                                                     |
|                                             | $\frac{\text{Output (decreases)}}{\text{Input (constant)}}$         |
| 3. Comprehensive decline.                   |                                                                     |
|                                             | $\frac{\text{Output (decreases)}}{\text{Input (increases)}}$        |
| 4. Declining efficiency.                    |                                                                     |
|                                             | $\frac{\text{Output (constant)}}{\text{Input (increases)}}$         |
| 5. Ineffective downsizing.                  |                                                                     |
|                                             | $\frac{\text{Output (decreases faster)}}{\text{Input (decreases)}}$ |

**Box 3: Five relationships for decreased productivity (Mistereck et al., 1992).**

In different industries and with different objectives for analysis, levels of nuance deepen for understanding productivity. Input-output definitions of productivity may also vary. Tangen (2005) lists 14 definitions, each with unique implications in context. Attempts have been made to distil the breadth of definitions—which vary in emphasis on efficiency, effectiveness, value and quality—into larger categories (Pritchard, 1995). Ghobadian & Husband (1990) suggest that the array of technical definitions be reduced to three broad conceptualisations. First, as a ‘technological’ concept, productivity can be understood narrowly as the output to input ratio for a productive process. Second, as an ‘engineering’ concept, productivity highlights the relationship between the actual and the potential output of a process given its inputs. Third, as an ‘economic’ concept, productivity is more concerned with the efficiency of resource allocation. The different conceptualisations of productivity correlate to different industries and different objectives. They highlight the importance of clarifying terminology and establishing basic understandings of the concept before conducting a study.

The purpose of productivity analysis may be to assess the productivity of an industry, an entire economy, or a single organisation. On one hand, production theory allows for numerous technical definitions of the concept. On the other, it provides a unifying frame of reference. Productivity is a concept that is endogenous to the object of analysis, and input-output investigation always focuses on systems functioning to transform initial resources into products and services with added value for their beneficiaries.

### 3.2.2 The Production Function

The production function is central to any productivity analysis. It is the abstraction of the production technology that transforms inputs into outputs (Gordon & Vaughan, 2011). Analytical objectives will determine whether investigators specify and explicate a function with definitive rules or whether investigators characterise the function as more of a black box. In production theory, clarifying assumptions about production functions is essential. The functions lend themselves to mathematical representations. The functions either (A) define the relationships between inputs and outputs or (B) are defined by patterns in observed data. In the former case, prior knowledge about an industry or *a priori* assumptions about production technology may be enough to specify functional characteristics and proceed with analysis in a deterministic fashion. Alternatively, when little is known about productive processes, functional rules may be inferred according to observed relationships between chosen inputs and outputs.

Conventional characterisations of production functions emanate largely from the field of economics. As covered later in the chapter, higher education productivity literature draws heavily from economic perspectives on the topic. It is thus prudent to build initial understandings with insights from this field. Although researchers and commentators debate whether foundational productivity principles are best attributed to the fields of engineering or operations analysis (Massy & Archer, 2018; Miles, Peterson, Miles, & Bement, 2018), key advances in production theory may be attributed to work from the field of economics.

A production function is generally denoted by the following:

$$Q = f(X) \quad [1]$$

In equation [1] the variable  $X$  represents an input or set of inputs to the production technology, and  $Q$  represents the output or set of outputs. Production theory literature, however, is largely concerned with ‘optimal’ productivity. That is, for a given set of inputs, production functions are often defined as generating the ‘maximum’ possible output. Thus,  $Q$  is generally understood to be the maximum amount of output that can be produced with the specified inputs by the production technology (Besanko et al., 2015).

Under assumptions that production functions and production technology serve to optimise production, a converse but logically equivalent interpretation of the function is often useful. The function may also be understood to ‘minimize’ input, given a certain level of output. The alternate interpretations of production technology are embodied in the principle of ‘duality’. The principle asserts that any input-output optimisation problem may be articulated in terms of a maximisation objective or its corresponding minimization objectives (Dogramaci & Färe, 2012). When financial margins are the prime concern—and when prices of both inputs and outputs are known—the duality principle is often extrapolated to investigate conditions for maximum revenue generation or for minimum expenditure. The foundational theory, however, applies to quantities of input and output, and value or quality could be defined and attributed to inputs or outputs by some other mechanism than price.

The purpose of establishing the baseline understanding of a production function in an optimal sense is largely prudent with respect to what organisations are set up to achieve with their inputs or outputs. Firms have been described as optimising their performance if they seek either to minimise price for customers or to maximise value by differentiating themselves among their competition (Porter, 2008). They can seek either to optimise efficient operations or to provide a uniquely line of products and services for customers. In minimising costs, there must be at least some minimum standard for acceptable output quality and amount produced. In maximising output value, there must likewise be a practical ceiling for what can be expended during production and development. In terms of input-output performance, Figure 4 synthesises performance possibilities discussed thus far and illustrates a spectrum of what is theoretically possible.

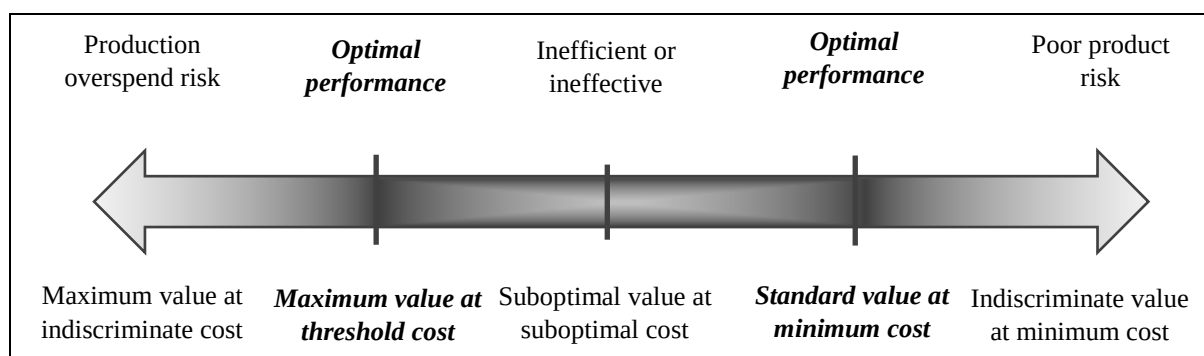


**Figure 4: Input-output performance spectrum**

From an economic perspective, when a single measure of performance—such as margin—can serve as an agreeable overarching proxy for success, then characterising optimal input-output performance is achievable with less ambiguity. Even when ‘value’ and ‘cost’ are more open to interpretation, however, the basic principles for what is optimal still apply. Designating appropriate and reliable measures and how much weight they should carry for decision-making becomes the primary challenge.

### 3.2.3 The Production Frontier

Defining a production function in terms of optimal performance helps to generate comparative scores of absolute productivity among peer organisations. It provides the ultimate standard of comparison within a given industry. If an optimal production function can be specified, then organisations may define the gap between their observed performance and optimal performance. For practical purposes, the production frontier is usually estimated with respect to the data of the highest performing organisations within an industry (Coelli, 2005). The definition of the frontier is either the maximum attainable output for a range of input levels or the minimum achievable input for a range of output levels.

An organisation's measured level of productivity compared to the industry frontier may be termed its productive efficiency, commonly referred to as 'technical efficiency' (Besanko et al., 2015). Technical efficiency is often designated as a ratio. The technical efficiency of Organisation A is defined as  $TE_A = \frac{P_A}{P_F}$ , where  $P_A$  is the measured productivity of Organisation A, and  $P_F$  is the industry's production frontier at corresponding levels of input or output. Technical efficiency in this sense can be thought of as a percentage of frontier efficiency, ranging from zero to one. (Kumbhakar & Lovell, 2003)

Figure 5 illustrates a production frontier curve in an industry where productivity is defined by one input and one output. Technical efficiency is illustrated with two example organisations that have different proximities to the production frontier. The technically inefficient example, Organisation A, is provided two reference points for improving productivity. Depending on optimisation objectives, it could seek to maintain its level of input and increase output—improving 'effectiveness'. Likewise, it could seek to maintain its level of output and decrease input—improving 'efficiency'.

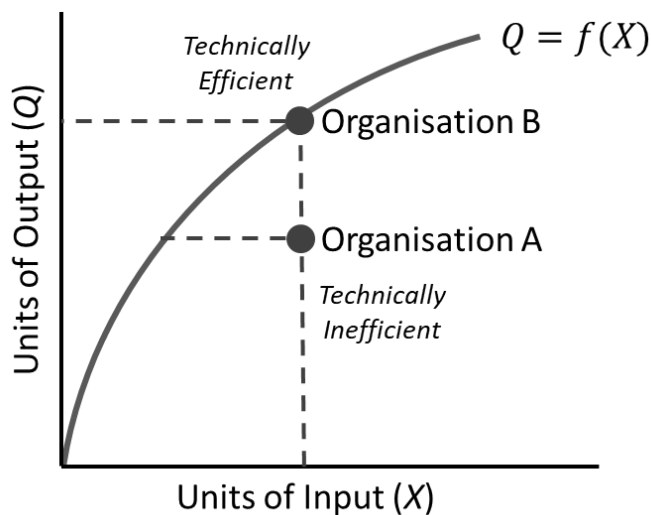


Figure 5: Single input and output production frontier curve

### 3.2.4 Productivity Change

Productivity change analysis provides an alternative to production frontier analysis. Two caveats around estimating production frontiers revolve around the empirical data needed to produce the estimates. First, noisy data can drastically affect frontier portrayals. Second, the

frontier is estimated only relative to institutions in the dataset. More fundamental or unobserved phenomena, or productivity insights from other industries are not reflected in frontiers estimated from single datasets (Dyson et al., 2001). Production frontier analysis can result in over-generalization about an industry, and the sophisticated techniques used to estimate optimal productivity may hide otherwise observable factors and dynamics that help with interpretation of measured results.

Many industries and organizations thus utilize a different class of production function, called productivity change functions. They are less definitive in showing relative efficiencies between institutions, but they can help reveal relationships between data elements and uncover key trends that drive overall behaviour. Calculations to estimating productivity change are also usually more straight forward. The Australian Bureau of Statistics (2012), the US Bureau of Labor Statistics (2007) and the OECD (2001) each endorse productivity change methods as standard protocol for measuring productivity across industries.

Productivity change estimation is less concerned about observed or theoretically optimal standards of comparison to gauge productivity. Rather, productivity change tracks performance trajectories and trends over time. Past performance and rates of progress are the standards for comparison. The phenomenon also has a precise mathematical interpretation. Productivity change may be associated with the derivative of the production function with respect to time (Caves et al., 1982). Estimating productivity change, however, does not first require specification of an absolute production function. The looser requirement for productivity change can be useful depending on analytical objectives. Levels of absolute productivity are ambiguous, but fewer assumptions need to be imposed upon data in question. Specific techniques are discussed in Section 3.4.

Analysis of productivity change is often referred to as ‘growth accounting’, and monitoring productivity growth can be associated with certain advantages and disadvantages (Antle & Capalbo, 1988; Mawson, Carlaw, & Mclellan, 2003). Growth accounting provides a way to track how institutions may respond to new policies, regulations, or changes in key work processes. It allows for portrayals of incremental progress and can serve as a lead indicator for broader performance growth or decline. As noted above, however, productivity change estimates do not readily give positions of institutions relative to a production frontier. Observed changes in a single institution’s productivity are thus open to question whether progress is due to industry-wide advancements or to innovations unique to that single institution. That is, change in productivity could be due to either the expansion of the production frontier or to a single organization moving closer to the frontier relative to its peers. If distinguishing between overall productivity change and change in technical efficiency needs to be addressed, then supplementary analysis of optimal production would need to take place. The analytical trade-offs, again, are tighter assumptions about the nature of the data and production, as well as characteristics of the institutions being analysed.

## 3.3 Productivity and Education

### 3.3.1 Departure from Foundational Economic Assumptions

Before unpacking the specifics of productivity estimation techniques, it is worth noting how input-output analysis in higher education has been influenced by broader insights from economics and how analysis has evolved over time. First, the prime currency of universities may be understood to be knowledge, not dollars (Coates, 2017b; Massy, 2016). The implication is that the production technology of most interest for analysis in the higher education sector is not akin to technological systems used to assemble tangible products to be sold on the market. Similarities do arise, however. Higher education institutions do use scarce resources to deliver products, services and achieve objectives. The direct outputs of a university may indeed represent only intermediaries in a pursuit of larger, longstanding mission objectives (P. Scott, 2006), but institutions still run on limited annual budgets, hire and provide for employees, acquire land and capital, and create value. Leaders, stakeholders and practitioners of higher education must still take part in the exercise of determining how an entire institution, certain departments, or individual activities should focus on optimising their value proposition to intended beneficiaries of education and research and to society at large. Indeed, stakeholders have the right to question whether public institutions efficiently and effectively use resources in pursuit of their missions.

HEIs cannot, however, be likened to traditional industrial firms in that they seek to minimise costs in pursuit of maximising profit. Universities are often state-run institutions and do not operate under strong assumptions of competition. Some HEI behaviour is steered by market forces—as discussed in the previous chapter with respect to rankings—but they do not behave as though they exclusively operate within ‘capital markets’. Institutional environments of universities have often been referred to as ‘quasi-markets’ (Marginson, 2013; Van Damme, 2015). Part of the reason for this is that higher education is still considered by many as a public good (Williams, 2016), and contemporary arguments affirm that their remaining as such would be positive for national economies and societies (Letizia, 2015).

Further to the point of how public universities serve to optimise their operations and value propositions to society, even base assumptions that they strive to minimise costs in pursuit of maximising notions of value or knowledge production is incorrect. The optimisation dynamics of how universities contribute to their missions is more complex. Public universities are known not to be cost minimizing entities. In fact, they have both monetary and legal incentives not to minimize costs. Hansmann (1980) explains that the primary difference between non-profit and for-profit enterprise is that the non-profit firm has legal limits on how surplus can be distributed. Any net earnings must not be paid out as profits to shareholders, but rather directed back into the organization to finance the services that the organization was established to provide. Hence, applying the cost minimizing dual to higher education productivity analysis would likely lead to both theoretically and practically invalid conclusions. James & Neuberger (1981) explain university priorities in terms of organizational utility or value functions, rather than in terms of standard production functions. Massy (2003, 2016) expands upon this notion, explaining that, while university expenditure serves as a constraint to mission value optimisation, the ‘duality’ principle of cost minimisation is not straight forward.

### 3.3.2 Human Capital Theory and Educational Rates of Return

In the absence of universal positivist theories of university productivity and behaviour, empirical input-output analysis has still generated insight for the higher education sector. Before delving into technical research opportunities, higher education productivity should be further framed in terms of the range of contemporary economic analyses that have been shaped enquiry within the sector. This section acknowledges seminal works in the study of economics and education that paved the way for current university productivity analysis, and (B) opens discussion for where further progress can be made.

The review is limited to work over the past half-century, beginning in the 1960s. Together, the works of Gary Becker and T. W. Schultz formed the foundations of ‘human capital’ study. Schultz (1960) defined human capital as the education obtained by an individual that “renders a productive service of value to the economy” (p. 571). Becker (1964) refers to human capital more generally as the resourcefulness of human beings, with specific reference to their wage-earning potential. Both researchers framed the concept of human capital as something in which to invest. Schultz (1961) and Becker (1962) represent separate studies, each entitled “Investment in Human Capital”. Both authors centre their discussions around investments in education and training—although healthcare and immigration are also mentioned as key investments to increase human capital within economic systems. With investment as the principal framing of the concept, questions naturally arise about the returns on such an investment. The study of human capital since its inception has thus been intertwined with studies about the rates of return on education—a precursor to technical input-output analysis of education institutions.

Becker (1960) uses a rates of return approach to study human capital by questioning the United States’ current level of investment in its own higher education system. The author speculated that many uncertainties existed for determining wage increases among those who attended college, as compared to those who have not. Reasons include (1) high variance in institutional types and quality, (2) the long period over which human capital accrues during time spent in university, and (3) the varying levels of qualification of students who are admitted to the institution. Schultz (1960) mirrors these concerns, stating that much work needed to be done to more accurately determine the true value of human capital gains attributable to educational attainment. Becker (1964), nevertheless, forms an empirical basis for larger claims. The author describes observations as reflecting the diminishing marginal product of time spent in the classroom. That is, earlier levels of education are associated with the greatest effects of increasing wage-earning potential.

The 1970s saw both expansion and added scepticism of educational rates of return analysis. Psacharopoulos (1972) extends analysis to international contexts, examining the return of education across 30 different countries at different levels of education. Welch (1970), however, raises questions about aggregate rates of return analyses for determining the productive value of education, noting discrepancies and peculiarities in rates of return data. Psacharopoulos (1970) also recognizes the issue by acknowledging the vast difference in observed earnings among individuals who have similar education credentials, and subsequently questions whether those wage differentials reflect education productivity differentials. The author devises an analysis using an explicit production function to unravel

aggregate results produced by rates of return computations. The analysis, however, is still finance-centric, using wages as the principal output of the function.

Spence (1973) further offers the foundations of an alternative to the human capital framing of education with the theory of market signalling. The theory challenges the idea that the responsible mechanism for increasing a college graduate's wage-earning potential is the education and training provided by the university. Rather, act of awarding of an accredited university qualification marks a discrete point in time when wage-earning potential increases because the degree itself is what signals competency to most employers, rather than the accumulation of years of training. Development of the theory has continued, and more recently Jaeger & Page (1996) show empirically the significant difference in wage-earning potential between individuals with and without a college degree, but who have accumulated the same amount of university training. Marginson (1993) also questions assumptions of human capital theory as they pertain to education, stating that such assumptions "have never been grounded empirically, yet they have found their way into common sense about education" (p. 31).

### **3.3.4 Toward Institutional Production Functions**

Overarching theories and aggregate studies were gradually accompanied by more technical studies, focusing on production and productivity at the intuitional level. (Johnson, 1978) discusses implications of examining higher education productivity at lower levels of analysis and of considering non-monetary factors, such as faculty time and credit hours as potential inputs and/or outputs to a production function. The author ultimately defaults to 'education valued added' and wage-earning potential as education's primary output, but incorporating more nuanced empirical factors into analysis, contributed to the field of study. By the late 1980s, (Monk, 1989) observes that education productivity research remained centred on large, structural factors, where findings presented only tenuous implications for nation or state-level policy. The author reopens discussion about the nature of lower-level education production functions. Although specific forms are left in question, the author asserts that many relevant education production functions could be defined to offer insight about core institutional operations and work processes (Monk, 1989).

By the 2000s, specifying institutional higher education production functions for administrative utility had taken hold. Research in higher education productivity began referring to a host of higher education production functions, rather than encompassing and absolute notions '*the* education production function' (Ehrenberg, 2004; J. Johnes, 2006; Mihaljevic, 2008). The precedent of bringing productivity analysis to the level of the institution presents new possibilities for evidenced-based decision-making, strategy, and resource allocation below macro and system levels. Production functions were conceived initially to examine output and input dynamics irrespective of finances. Although common application of production functions in other industries gravitate toward representations of revenue and profit—and although early productivity studies in education primarily explored rates of return—the higher education industry has begun utilizing more nuanced methods that attempt to speak more directly to the foundations of education key work processes.

### 3.4 Taxonomy of Productivity Measurement Models

#### 3.4.1 A Universe of Measurement Models

This section presents an original taxonomy for identifying and classifying different types of productivity measurement models. The taxonomy is necessary to help reveal and define gaps where research is needed to advance the practice of measurement in higher education. Most research on university productivity has occurred in the economics of education literature and centres on particular mathematical techniques to produce productivity ratios from given data (De Witte & Sneyers, 2017; G. Johnes & Johnes, 2009; Szuwarzyński, 2019). Economic paradigms and associated assumptions, however, necessarily recognise a narrow set of model components to be manipulated. The taxonomy in this section is presented in hierarchical fashion. It starts with foundational and often implicit considerations for designating a model. Five productivity modelling dimensions are discussed.

Figure 6 shows the taxonomy. The modelling dimensions include: (1) time differential, (2) input characteristic, (3) output characteristic, (4) production complexity, and (5) data aggregation method. Decisions about the time differential encompass foundational analytical objectives about whether trends and dynamics over a period are of prime interest or whether absolute and comparative productivity scores at single points in time are of prime interest. The time differential refers to measuring either absolute productivity or productivity change, as discussed in section 3.2 above. The input characteristic refers to the types and nature of input data elements within the productivity framework. The output characteristic refers to the types and nature of output data elements in the framework. Production complexity refers to the nature of the production technology being analysed. The data aggregation method refers to the mathematic techniques specified to calculate a productivity ratio given the input-output framework and complexity production processes.

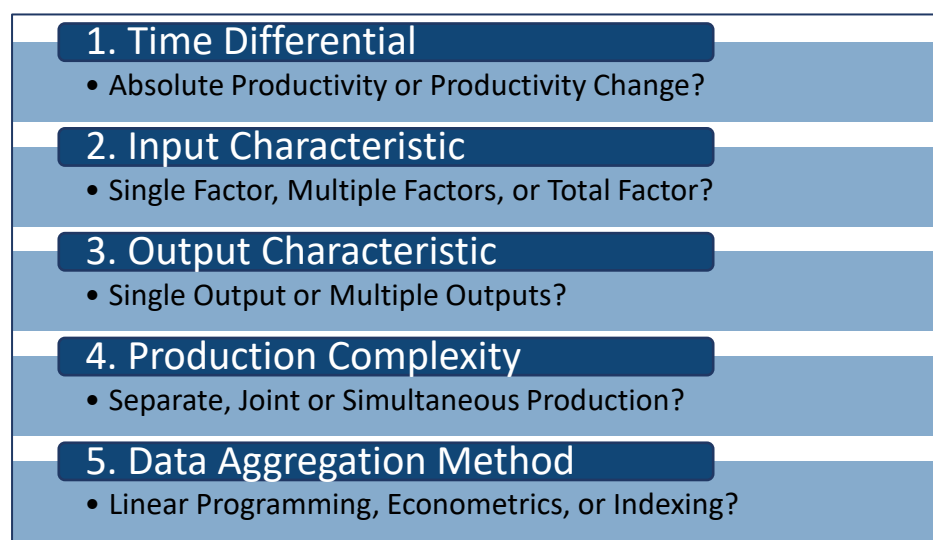


Figure 6: Production function taxonomy

### 3.4.2 Input and Output Characteristics

The taxonomy discussion proceeds to clarify input and output characteristics. Key factors associated with a measurement model's time differential—absolute productivity and productivity change—are discussed above in section 3.2. The prime input characteristic is whether the model measures single factor productivity (SFP), multifactor productivity (MFP) or total factor productivity (TFP). The designation relates to whether outputs in the model are explained by a single input, more than one input, or all conceivable inputs used during the production process. The OECD (2001) recognises that TFP and MFP are often used interchangeably, but in this taxonomy treats them as distinct concepts. Examples and explanations are provided in the following paragraphs.

The designation SFP refers to a measure of productivity that considers one input to production. The input is most often a measure of labour, for example numbers of employees or total hours worked (Besanko et al., 2015). Traditionally, human labour has been the chief essential input to productive processes. In many industries, as in education, labour expenses further represent a large, if not the largest proportion of expenses toward production (Miles et al., 2018). The designation MFP refers to a measure of productivity that considers multiple types of productive inputs. A natural expansion from tracking only labour inputs is to track both labour and capital (Besanko et al., 2015). In contemporary higher education institutional studies, diverse combinations of inputs are tracked. Wolszczak-derlacz (2017) considers staff, student, and revenue data as inputs, whereas Tran and Crawford (2017) consider staff, campus area, and expenditure data.

The designation TFP refers to a systematic technique that attempts to aggregate and capture all input factors of production (Antle & Capalbo, 1988). The OECD (2001) summarises the factors as capital, labour and intermediate inputs, often abbreviated KLEMS, or capital, labour, energy, materials, and services. Sullivan et al. (2012) specified methods for tracking TFP in higher education, and the methods have since been replicated or adapted in (Miles et al., 2018; Moore, Coates, & Croucher, 2018a).

The prime output characteristic refers to whether only one output or multiple types of outputs are being measured in a productivity ratio. If one type of output is measured, it could be the quantity of 'widgets' produced, or it could be a financial indicator measured in dollars, such as value-add (Bairam, 1994; Tangen, 2005). When outputs are being sold on the market, incorporating their prices after sale not only helps firms track only the outputs that are bringing utility to customers, but also sales prices help with aggregating different types of products and services into a single monetary figure related to revenue. Even if multiple types of tangible outputs are represented under monetary figures, this taxonomy treats all financial data elements as a single type of output.

A model with the designation of multiple output types incorporates at least two outputs measured on different scales. The units of the outputs must be different. Attempting to capture multiple output types in higher education productivity measurement is common practice. Because key academic outputs are not often sold on the market, they must be tracked according to more inherent characteristics, such as numbers of graduates or research publications. While Yaohua et al. (2018) examine numbers of students enrolled and numbers of scholarly publications, Johnes (2006) examines numbers of degree completions and research-related income.

### 3.4.4 Production Complexity

The next component of the taxonomy is production complexity. The specification of this model dimension depends upon recognising the nature of the production technology that transforms inputs into outputs. Mishra (2010) summarises two overarching types of production functions, single product functions and joint product functions. Single product technologies take inputs and result in one type of output, whereas joint product technologies result in multiple types of ‘co-products’. The specification has both practical and technical implications. Single output production functions signify simpler and more straight forward production processes, and joint functions signify more complex processes. The recognition of either type has mathematical implications for the measurement model.

In higher education, where institutions often produce more than one types of output, it is possible to examine certain processes separately, in terms of single-product functions, or examine them together in terms of joint production functions. Chizmar & Zak (1983) provide more precise definitions of three distinct production technology representations for education production processes. The representations include: (1) separate production technologies, (2) joint production technologies, and (3) simultaneous production technologies.

Separate production technologies incorporate the most straightforward generalization of the standard production function presented in section 3.2.2. Although the number of institutional outputs might be numerous, for simplicity, we limit the definition to two distinct outputs of production  $Q_1$  and  $Q_2$  and two input vectors  $\mathbf{X}_1$  and  $\mathbf{X}_2$ . The mathematical definition is provided below.

#### Separate Production [2]

$$\begin{aligned} Q_1 &= f(\mathbf{X}_1) \\ Q_2 &= g(\mathbf{X}_2) \end{aligned}$$

The separate production system of equations [2] represents two distinct production technologies. Conceptually and practically they operate independent of one another and could conceivable run on different time frames. Fizel, Gustafson, & Hadley (1999) add to the definition of separate production technologies, noting that there exists no causal or dependent relationship between the outputs of the separate processes. Further, separate production functions require understanding the different amounts of input allocation for each output.

Joint production technology requires a more complex generalization of the standard production function. First, there is no requirement to know specific amounts or types of inputs allocated toward the production of each type of outputs. Further, joint production implies a consistent relationship between the different types of outputs, and thus requires the specification of an additional function to relate and aggregate the co-products. The mathematical depiction of a joint production technology for two co-products  $Q_1$  and  $Q_2$  with a vector of inputs  $\mathbf{X}$  is provided below.

#### Joint Production [3]

$$g(Q_1, Q_2) = f(\mathbf{X})$$

Joint production [3] is the most common representation of higher education production technology in the scholarly literature. The concept is discussed further in the next section covering data aggregation methods. Examples of specific studies that employ the technique are then provided in section 3.4.

Simultaneous production technologies exhibit the highest level of complexity and the most interconnected relationships between inputs and outputs. Specification of this type of function requires the most nuanced understanding of the production technology being examined. For simultaneous production, the different types of outputs have separate production processes, but they are interrelated and co-dependent. The output of one process is necessary for the output of the other process, and *vice versa*. The mathematical specification helps to clarify the relationships. Consider again two production technologies for two outputs  $Q_1$  and  $Q_2$  and two separate sets of inputs  $\mathbf{X}_1$  and  $\mathbf{X}_2$ . The definition is provided below.

#### Simultaneous Production [4]

$$\begin{aligned} Q_1 &= f(Q_2, \mathbf{X}_1) \\ Q_2 &= g(Q_1, \mathbf{X}_2) \end{aligned}$$

Fizel et al. (1999) summarizes simultaneous production [4] as two separate production processes where there is some causal and dynamic relationship between the outputs. W. Becker & Walstad (2012) discuss possibilities that simultaneous functions may serve as the most appropriate conceptual representations for education production contexts. Simultaneous production of education and research outputs would certainly be the most representative mathematical abstraction of the higher education ‘teaching-research nexus’. Yet due to mathematical complexity and empirical limitations, simultaneous production is the least common specification of production technology in the scholarly literature.

#### 3.4.5 Data Aggregation Method

The final dimension of the productivity measurement model taxonomy is the data aggregation method. Based on specifications for the first four model dimensions discussed above, a variety of calculation techniques could be chosen to generate a ratio of output to input. Each has strengths and weaknesses. Singh et al. (2000) summarize three umbrella methods to estimate productivity: econometric techniques, linear programming, and index measures. The following paragraphs explain each of the three main methods.

The first data aggregation method discussed is econometric. Econometric techniques typically estimate production functions using regression analysis. Based on either prior knowledge about an industry or relevant production technology, a deterministic functional form with known parameters relating inputs and outputs should be set. Regression analysis is usually performed with respect to the specified function to generate a production frontier, and the formulas generally require pricing data on inputs and outputs. The technique allows for deep investigation of input and output dynamics, revealing economic indicators such as returns to scale, scale elasticities, and marginal costs (Bairam, 1994). If little inherent knowledge about the production technology is available, flexible functional forms are possible, but such techniques require larger amounts of data (Singh et al., 2000).

Stochastic frontier analysis (SFA) is the most common econometric technique for analysing productivity in higher education. The method uses a stochastic error term to explain variation in the relationships between the variables and to help estimate a production frontier when data do not naturally fit regular patterns (Kumbhakar & Lovell, 2003). For universities, the function being estimated is usually a multi-product cost function, where the dependent variable is cost (the input), and the independent variables are institutional outputs (Izadi, Johnes, Oskrochi, & Crouchley, 2002). Titus and Eagan (2016) explain that SFA has been less common for analysis in higher education because of its restrictions in providing for both multiple inputs and multiple outputs. In Australia, the most prominent applications of SFA appear in professional reports, such as Deloitte (2016).

The next data aggregation method is linear programming. Linear programming is a non-parametric technique used to estimate a production frontier without first needing to designate a parameterised functional form. Production frontiers are constructed in a piece-wise manner according to precise characteristics of the dataset. Consequently, the technique produces no consistent rules for understanding relationship dynamics between inputs and outputs (J. Johnes, 2006). Alternatively, non-parametric analyses hold the potential to identify a production frontier that is highly representative of the dataset (Ray, 2004). Because residuals are not being minimised with respect to preset function, the frontier may naturally follow the data in irregular patterns. Conclusions to be made from such analyses, however, are specific to the dataset and less generalizable outside specific contexts. Non-parametric techniques hold the potential to reveal patterns in complex multi-input, multi-output data, and in the absence of specific knowledge about production technology, they can provide a viable solution for inquiry. Further pricing information about inputs and outputs is not required, and wider variety of data elements is thus possible to analyse.

DEA is the most common linear programming technique for analysing higher education data (De Witte & Sneyers, 2017). As noted above, a strength of DEA for university productivity analysis is its ability to aggregate multiple inputs and multiple outputs without the need to designate a parameterised functional form for relating data elements (Cooper, Seiford, & Tone, 2007). The objective of DEA's linear programming in aggregating data is to maximise each institution's productivity portrayal by determining optimal value weights for each data element in the model. The result is that each institution is portrayed in its most favourable light with respect to all other institutions (Coelli, 2005). Researchers are not required to manually assign value weights to inputs and outputs in the aggregation process, which may be viewed as either a strength or a weakness of the technique. On one hand, researchers may avoid including their own value judgments or biases in assigning value to non-priced academic outputs (Cooper, Seiford, and Tone 2007). However, this does not mean that value judgments are not being conveyed. The task is relegated to the computer, and DEA derived value weights for education and research outputs emerge from characteristics of the dataset and not from broader institutional context. Further, the value weights are often not reported (Abbott & Doucouliagos, 2003; Carrington et al., 2005; Moradi-Motlagh et al., 2016), and the differential emphasis given to education or to research in each institution is not clear. Lee and Worthington (2016) explain that "the rigidity of standard DEA treats the production process as a 'black box' in that it simply transforms inputs into outputs and neglects any possible intervening processes" (p.26).

The final data aggregation method is indexing. Similar to linear programming, index numbers are useful under general circumstances when advanced knowledge about the nature of production technology is incomplete (Caves et al., 1982). Index methods are flexible in terms of the types of data elements that can be used, and options exist for incorporating and adjusting for price information. The function of index numbers is to generate individual productivity indicators corresponding to specific and observed points in an empirical set of data. Unless paired with additional methods, the indexing approach does not compute solutions to internal problems, such as fitting data to curves with regression or optimising estimates with mathematical programming. The result is that typical index methods allow only for measurement of productivity change because the techniques are not designed to fill gaps in observed data with predictions of a production frontier. When measuring only productivity change without reliable base productivity estimates, however, care must be taken when interpreting results because distortions in the data become possible (Singh et al., 2000).

An emerging index method in higher education is the Törnqvist index (TI). The TI technique computes productivity change per institution by aggregating year-on-year percent change ratios of data elements. The TI formula divides changes in a composite output variable by changes in a composite input variable. It is deemed a superlative index because of its correspondence to flexible functional forms in econometrics, allowing for deeper analysis of input-output relationships and of additional economic concepts if data are adequate (Caves et al., 1982). Because of this characteristic, and because TI estimates reflect observed data with fewer back-end calculations, the index has been advanced for TFP measurement across industries by the Australian Bureau of Statistics (2012), the US Bureau of Labor Statistics (2007) and the OECD (2001).

### **3.5 Contemporary Models and Gaps in Knowledge**

#### **3.5.1 *Ad Hoc* Nature of Current Approaches**

There has been relatively little consistency in university productivity analysis across higher education internationally. Even if linear programming has shown to be the most popular data aggregation method, (1) different combinations of the other modelling aspects allow for considerable diversity, and (2) estimates produced from linear programming render results minimally generalisable beyond the data set under investigation. These points together with diverse contextual factors and stakeholder interests associated with university performance leave the study of university productivity in an *ad hoc* state of inquiry. Generalising about higher education productivity may always be difficult given idiosyncratic tendencies that depending on a host of national and local circumstances, given the nature of its unpriced outputs, and given the difficulty in accurately tracking learning and research outcomes. The study of higher education productivity is nevertheless still open for many potential advancements that could assist in both unifying and grounding methods, assumptions, and objectives for analysis.

To illustrate the inconsistency of university productivity analysis, Table 5 provides a sample of 20 productivity studies within higher education since the mid-2000s. The table provides a mapping of the studies according to measured output indicators and to the data aggregation method employed. Across all studies, only two models use both a common set of outputs and aggregation method. With so much diversity, the relative value of different methods in context can be difficult to judge.

The range of potential outputs shown in Table 5 is indicative of the multiple paradigms for understanding productivity discussed in section 3.2.1. Data selection within a study serves as the expression of the operative analytical paradigm and what researchers and stakeholders have deemed important for representing performance in context. For example, Nazarko (2014) considers student achievements as educational outputs by tracking scholarship recipients and study abroad participants. Olariu & Brad (2017) consider enrolment numbers and government revenue. Each speaks to a different realm of performance and priorities, but both do so under the guise of productivity. Each study also computes results using DEA. More consensus exists around tracking research output via academic publications, but specific measures and metrics vary.

**Table 5: Productivity models in higher education**

| Indicator         | Variable                   | Calculation Technique |   |   |   |   |   |   |   |   |   |              |   |   |   |          |   |   |   |
|-------------------|----------------------------|-----------------------|---|---|---|---|---|---|---|---|---|--------------|---|---|---|----------|---|---|---|
|                   |                            | Linear Programming    |   |   |   |   |   |   |   |   |   | Econometrics |   |   |   | Indexing |   |   |   |
| Education Outputs | Student Load               |                       |   |   |   | X |   |   |   | X |   |              |   |   | X | X        |   |   |   |
|                   | Student Enrolment          | X                     | X | X | X |   |   |   |   | X |   |              |   |   |   |          | X |   |   |
|                   | Student Achievements       |                       |   |   |   |   |   |   |   | X |   |              |   |   |   |          | X |   |   |
|                   | Degree completions         |                       |   |   | X | X |   | X |   |   | X | X            | X |   |   | X        | X |   |   |
|                   | Adjusted Credit Hours      |                       |   |   |   |   |   |   |   |   |   |              |   |   |   |          |   | X | X |
|                   | Exceeded or passed credits |                       |   |   |   |   |   |   |   |   |   |              | X |   |   |          |   |   |   |
|                   | Graduate employment        |                       |   |   |   |   |   | X | X |   |   |              |   |   |   |          | X |   |   |
|                   | Revenue from Government    |                       | X |   |   | X |   |   |   |   |   |              |   |   |   |          |   |   |   |
|                   | Revenue from Student Fees  |                       |   |   |   | X |   |   |   |   |   |              |   |   |   |          |   |   |   |
| Research Outputs  | Publications               | X                     |   |   | X | X | X |   | X |   | X |              |   |   | X |          | X |   | X |
|                   | Citations                  | X                     |   |   |   |   |   |   |   |   |   |              |   |   |   |          | X |   |   |
|                   | Institutional Rankings     |                       |   |   |   |   |   |   |   | X |   |              |   |   |   |          |   |   |   |
|                   | Research Degrees Awarded   |                       |   |   |   |   |   |   |   |   |   |              |   |   |   |          | X |   | X |
|                   | Patents                    |                       |   |   |   |   |   |   |   |   |   |              |   |   |   |          | X |   |   |
|                   | Research Income            | X                     |   |   | X |   |   |   |   |   | X | X            | X | X |   |          | X | X | X |
| Misc.             | Reputation                 | X                     |   |   |   |   |   |   |   |   |   |              |   |   |   |          |   |   |   |

*Yanolua et al. (2018)*  
*Olariu & Brad (2017)*  
*Tran et al. (2017)*  
*Wokczak-Derbez (2017)*  
*Moradi-Motlagh et al. (2016)*  
*Arijmandi et al. (2016)*  
*Avilez-Sacoto (2015)*  
*Nazarko & Suparamasuk (2014)*  
*Klimpp & Zdzewski (2013)*  
*Worthington & Lee (2008)*  
*Johnes (2006)*  
*Carrington et al. (2005)*  
*Abbot & Doucouliagos (2003)*  
*Salas-Velasco (2018)*  
*Titus & Fagan (2016)*  
*Agaski & Johnes (2015)*  
*Izadi et al. (2002)*  
*Coates (2017)*  
*Miles et al. (2018)*  
*Moore et al. (2018)*

The frequent use of linear programming methods for higher education productivity analysis illuminates both strengths and weaknesses of the state of the art of measurement. Non-parametric, linear programming methods are the most mathematically intensive and have been designed with the objective to treat complex datasets. On one hand, their use represents an acknowledgement of the unique complexities of university productivity analysis, and an admission that simpler, deterministic models would likely treat nuanced data with too heavy a

hand and generate questionable results. They also illustrate attempts to develop higher-education-specific models, rather than importing template models from other industries.

Conversely, employing the most complex and most difficult-to-decompose models introduces risks for misinterpreting data. Generating *ad hoc* production functions before exploring more fundamental trends and foundational theories about behaviours observed in the data may represent forgone insight. DEA methods allow for levels of precision in representing common productivity concepts, such as technical efficiency and movement of efficiency frontiers—which less complex techniques cannot achieve. Prioritising technocratic portrayals of data, however, over external validity and contextual suitability runs risks of generating precise conclusions about the wrong phenomena.

Carrington, Coelli, & Rao (2005) lament “the dearth of appropriate comprehensive productivity measures for the university sector” (p. 147). The authors explain that universities’ non-for-profit nature and their multiple, unpriced outputs create limitations. Under the circumstances more common and preferred methods, such as the Törnqvist TFP index, are unavailable. The authors explain that, without price information on outputs, TI’s value weighting procedures for output data element aggregation cannot be achieved empirically. This is the principal reason for employing DEA methods, which uses “shadow price information that is implicit in the shape of the estimated DEA production frontiers” (Carrington et al., 2005, p. 147). Sullivan et al. (2012), however, establish theoretical and practical foundations for using the TI in higher education and advance a methodology to incorporate more than one educational output. Work still needs to be done for TI methods to capture research output in conjunction with education output, but opportunities for doing so are introduced and tested in (Moore, Coates, & Croucher, 2018b; Moore et al., 2018a).

### **3.5.2 Specifying Gaps and Opportunities**

Insights from the literature review—including insights from Table 5—are corroborated by conclusions made in De Witte & Sneyers' (2017) literature review on efficiency in education. Together they reveal broad and specific gaps in knowledge and opportunities for further research. The taxonomy introduced in section 3.4.1 further specifies five unique dimensions of any university productivity measurement model, and the taxonomy can aid in explicating where new technical work is needed. Although the overarching research questions of this dissertation transcend technical aspects of measurement, this section reveals major areas for advancement of productivity measurement tools. Section 3.6 below synthesises broader contextual and conceptual information to provide a rationale for three more overarching research questions intended to guide subsequent work in this PhD project. First, though, clear gaps are identified, which if filled, complement objectives to generate broader, more practical insight for higher education policy and practice in Australia.

Broad opportunities for productivity measurement in higher education revolve around incorporating insight about measurement issues from other fields and integrating knowledge about the context of higher education provision. As discussed above, the current mathematically intensive measurement techniques deployed on higher education data lend themselves to an inward focus on measurement model components and limitless tweaking of mathematical procedures. In higher education, “the efficiency literature pays currently a significant effort to optimize its own models. In doing so, it focusses on minor

methodological details and neglects the current important issues in related fields” (De Witte & Sneyers, 2017, p. 356). Further, the singular focus on linear programming methods limits opportunities for insight about institutional functioning and performance. With reference to all three umbrella methods used across industries—econometrics, linear programming, and indexing—input-output formulas can sometimes be deceiving in representing overall performance (Singh et al., 2000). The author urges that productivity measures be designed and tailored to reflect the operational strategy of the firm or institution being assessed.

A major technical objective of this dissertation is to advance the development and use of indexing methods for higher education productivity measurement. In providing a rationale and approach for doing so, this section explicating gaps in knowledge and research opportunities with respect to the productivity measurement taxonomy from the previous section, and with respect the literature review performed for this dissertation, which serves as the foundation for the content of the current and previous chapters. Specific studies from the literature are cited to exemplify gaps in knowledge and the urgency for more research in this field.

Figure 7 illustrates different possible combinations of model characteristics with modelling decisions concerning the first two dimensions in the taxonomy, time differential and input characteristic. Based on the literature review, absolute productivity frontier analysis using MFP measures dominates the field. Clear inroads for further research and development lie with TFP measures of productivity change. There are examples of single factor productivity analysis in the literature, and there may be cause to develop these methods further, but the multi-input multi-output nature of the industry lends itself to methods that provide for generating both composite input and output indicators.

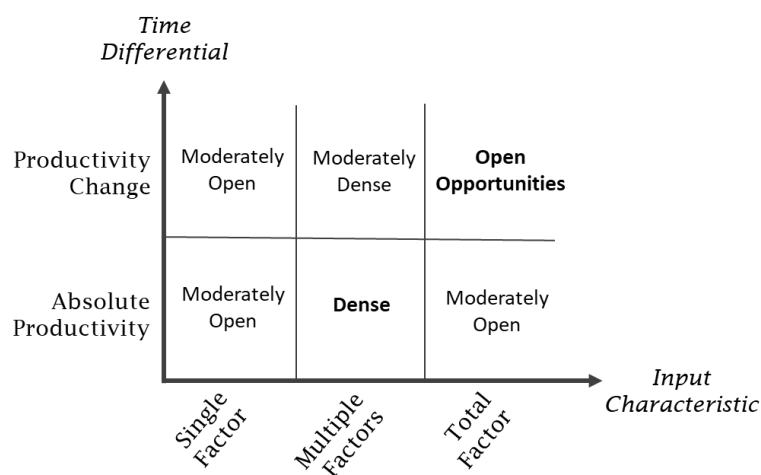


**Figure 7: Productivity models used in this research**

The moderately dense literature on MFP productivity change bares noting and characteristics of such studies serve as prime justification for further advancing the literature on TFP change measures. The bulk of studies that measure productivity change in the higher education produce results that derive from absolute productivity frontier estimates at different points in time. The implication is that productivity change estimates are generated from methods designed to estimate production frontiers and absolute productivity. In principle, this should not raise issues, however, any concerns associated with the non-parametric linear programming methods used to estimate production frontiers would carry over to estimates of

productivity change. Alternatively, broader productivity measurement literature advances numerous techniques designed to generate reliable estimates of productivity change. Among the most common techniques are those designed to estimate TFP change. Sullivan et al. (2012) further explicate how such techniques can be applied to higher education.

Figure 8 shows the difference between the approaches of most prior work on higher education productivity and new potential opportunities to make contributions to the field. Differences are expressed with respect to the dimensions of output characteristic, production complexity, and data aggregation method. Most studies use assumptions of joint production to model multiple outputs with linear programming techniques. A prime contribution of this dissertation is to explore the use of indexing methods. Indexing methods can be refined to handle multiple outputs, as well as joint and separate education and research productivity. There are significant opportunities to contribute to the field using indexing techniques because all decisions about data aggregation must be made by the researcher. The process may be cumbersome, but the use of index measures opens up the black box of productivity measurement. Index measurement with multiple inputs and multiple outputs is not conducive for producing results in an expedient manner. Numerous scholarly and practical opportunities exist, however, for taking the time to conduct a measurement exercise using this method.

Recall from the

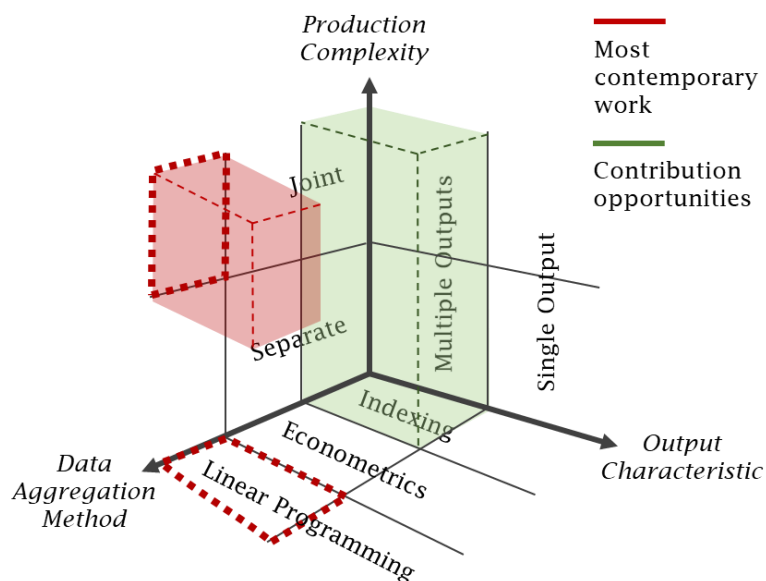


**Figure 8: Space for scholarly contribution**

### 3.6 Research Question Formulation

This chapter characterised the state of the art of productivity measurement as it applies to higher education. While dozens of studies measuring higher education productivity can be identified, the field of study is relatively unestablished. The technical and conceptual background provided in this chapter has illuminated where research opportunities lie, and

where targeted questions need to be posed to begin filling gaps in knowledge. The review and synthesis of the literature lays a foundation for making technical and conceptual advances, as well as to incorporate key contextual information into the practice of in the Australian context.

The underlying motivation for this research is to characterise the productivity of the Australian higher education system. Public data on Australian higher education is extensive, and large datasets invite complex measurement techniques. Many options exist for synthesising, analysing and summarising the data, but the aim of this research is to generate university productivity insights that maximize value and utility for institutional leaders and stakeholders—not to optimise mathematical models with respect to collected data. Major questions about what can and should be determined about university productivity in Australia are thus left open at the outset of the project. This research does not assume that the most useful knowledge about university productivity lies in knowing whether productivity is increasing or decreasing, whether some universities are more productive than others, or proximity to an efficiency frontier. Accordingly, strict hypothesis testing is not part of the methodology. Important insights to be inferred from university productivity analysis should be guided by key stakeholders, and this research leaves options and directions for quantitative analysis open as qualitative insight emerges.

To form a useful characterisation of Australian higher education productivity three essential questions must be posed and explored. The next chapter on research design explains how the research question are explored iteratively and simultaneously throughout the project, but conceptually it is helpful to frame them sequentially. The first question is stated as follows:

**Q1: How can different approaches to understanding productivity in higher education be characterised?**

The first question has been explored to some extent over the course of the literature review. Most academic literature that frames broad understandings of productivity in higher education, however, does so in technocratic light and is discussed by researchers, practitioners and experts directly studying the phenomenon or measuring the phenomenon in some applied sense. This research seeks to form broader conceptual notions of productivity in higher education as understood by a range of diverse stakeholders with no direct connection to the formal study of higher education productivity. The intention is to generate more encompassing understandings of the concept with respect to the specific Australian context and with the possibility to generalise more broadly. The intention is to better align potential productivity measurement tools and results with the widest possible range of stakeholder interests. The intention leads to the second research question:

**Q2: What are the characteristics of a measurement model appropriate for investigating productivity of Australian universities?**

With broader understandings of higher education productivity, and with insight from a wider range of stakeholders, this question confronts the problem of using such insight to develop the most appropriate measurement tools for the Australian context. Technical literature reviewed in this chapter highlights where opportunities lie. Addressing the question, thus, involves combining insight from what has already been learned in the technical scholarly realm with local, contextual knowledge about higher education policy, practice and provision in the

Australian system. With an appropriate and contextualised measurement tool, the third question can be addressed:

**Q3: How can the productivity of Australian universities be characterised?**

As mentioned above, the third research question is intentionally broad and open. The question is intended to showcase the potential insight the measurement model can provide when its results are used to inform a productivity assessment of the Australian higher education system. No expectations are imposed upon productivity findings because ideas about the most important concepts and phenomena to represent in the model and in the assessment of Australian university productivity are expected to unfold in process of addressing the first two questions. The next chapter describes the formal methods chosen to address these questions, and they leave open the possibility for making unconventional conclusions about university productivity that does not revolve around conventional, economics-influenced frontier efficiency measurement.

## 3.7 Summary

### 3.7.1 Focus and Contributions

Chapter 3 has added technical evidence to support the thesis of this research. University productivity measurement can and should be improved to allow for authentic and reliable performance assessment of institutions in context. The literature on university productivity measurement confirms technical difficulties in generating productivity estimates from institutional performance data, but it also offers solutions. There exists a substantial literature on university productivity and efficiency measurement. Contemporary research, however, often seeks to optimise mathematical models for statistical precision in relation to available data. Higher education productivity measurement studies are thus isolated and *ad hoc*, with few similarities beyond a preferred calculation technique, called DEA. This chapter has argued that advancing measurement means considering an alternative method, called indexing. Using this method—and the expanded analytical frame required to employ it—could widen the scope for insight of productivity measurement in higher education.

Crucially, this chapter identified research opportunities for advancing measurement based on the broad improvement areas identified in Chapter 2. This chapter presented an original taxonomy that explicates gaps and opportunities for improving measurement practice in higher education. The taxonomy illustrates how different measurement models have different scopes for insight. The space for further research on university productivity measurement is the parameterised by the taxonomy and other background provided in the chapter. Finally, The parameterisation sparked three research questions to guide the remainder of the project.

### 3.7.2 Considerations and Limitations Arising

The chapter characterised inevitable technical problems for higher education productivity measurement; namely, the challenge of combining different types of performance data into a single, aggregate productivity indicator. The problems, however, are not intractable with sufficient nuance in analysis. Simple productivity calculation techniques are not an option in higher education because of universities' unpriced inputs and outputs. Data elements cannot be easily normalised and aggregated by price to give estimates of total monetary value in and total monetary value out. DEA solves the data aggregation problem. The linear programming methods of DEA provide an empirical rationale for how different types of inputs and outputs can be aggregated. While the aggregation algorithms of DEA are empirically based, however, they rely exclusively on characteristics of the input-output data under consideration. Results are thus not guaranteed to reflect contextual circumstances unrepresented by the study's input-output data framework.

Indexing methods can also aggregate multiple inputs and outputs, but no automatic aggregation takes place. Parameters for aggregation must be set by the researcher and are

likely based on outside knowledge not contained in the study's input-output framework. This has been described as a limitation of the indexing method because it allows for researcher bias to influence measured results. There is a trade-off, however, because DEA results are also biased. DEA bias, however, is just the result of an indiscriminate, automated computer algorithm—rather than that of a human being. The use of indexing methods inevitably requires multiple layers of additional justification for the treatment of individual data elements in calculation. But the conduct of a meticulous measurement exercise is where the value lies with an indexing measurement approach. DEA can give statistically robust estimates for a dataset in an expeditious manner. Indexing, however, through the embrace of researcher judgement and external input, may sacrifice precision but also may increase accuracy and contextual suitability.

## CHAPTER 4: RESEARCH FRAMEWORK AND METHODOLOGY

*It's only been the last few years that the sector has been looking at some big intractable problems through the lens of design thinking. We're saying, give us the thorniest issue in policy, practice or procedure, and we'll actually review it.*

–Study Participant KS.9  
Academic Program Director

### **Abstract:**

This chapter provides a rationale for the research methodology and argues why the chosen methods are appropriate for addressing the study's research questions. The methods are chosen to facilitate the design of a measurement model for assessing university productivity. The design processes is intended to generate learnings that contribute to knowledge, theory and practice. The chapter explains how a design science research methodology is best suited for the project. The project comprises eight stages. The chapter summarises each stage to elucidate all research activities undertaken throughout the PhD program. Key activities include adapting and generalising a measurement model introduced by the United States National Research Council (NRC), testing the model across international contexts, synthesising interview responses from experts and key stakeholders to refine the model, and analysing measurement results of the final model to assess Australian university productivity.

### **Key Words:**

Measurement, models, evaluation, assessment, design science

## **4.1 Introduction**

The methods in this research project are chosen to unlock conceptual, technical and practical knowledge on higher education productivity and its measurement. Key activities include large-scale collection of secondary institutional performance data on all Australian universities and the development of a measurement model to provide productivity change estimates from the data collected. Methods further include structured interviews with Australian experts and stakeholders to clarify Australian higher education productivity issues and to inform the interpretation of results generated by the measurement model. The project concludes with a final comprehensive data analysis and a productivity assessment of Australian universities. The research methodology that unites all research activities is design science research.

Design science ensures a consistent and cohesive structure of enquiry for entertaining conceptual, technical and practical research questions. The gaps in knowledge identified in the previous chapter suggest that the design process of a new productivity measurement model can serve as a vehicle for investigation. The output of the design process—the new measurement tool—addresses Question 2, which regards suitable characteristics for a measurement model. Arriving at those characteristics, however, requires addressing Question 1, which regards identifying various understandings of higher education productivity. The final measurement model can then be used on university performance data to address Question 3, which regards characterising the productivity in Australian higher education institutions.

The content of the chapter includes a full explanation of the principles, practice and procedures of design science, and how it is used to organise the research in this dissertation. The chapter then details the research project design. Key contents include: (1) details about how practitioners from nine Asia-Pacific countries were able to test and provide feedback on the model under design in this research, (2) qualitative methods used for 20 Australian key stakeholder and expert interviews, (3) quantitative methods for the final data analysis, and (4) an explanation for how learnings from across the study are synthesised to codify new knowledge and to reveal implications for policy and practice.

## **4.2 Research Methodology**

### **4.2.1 Design Science Research**

This study employs a design science research methodology to provide structure for an investigation that seeks to produce both qualitative and quantitative findings. Design-science

has emerged to unite the technical and the human-centric aspects of research problems. Plattner, Meinel, & Leifer (2016) advance the idea that design thinking creates a bridge between the qualitative, context-dependent aspects of a problem and the scientific, context-independent aspects of a problem. The methodology is more than a set of steps for product design or market innovation and has matured via multi-disciplinary use in addressing complex, socio-technical issues (Rowe, 1991). Peffers et al. (2008) describes the design science methodology in terms of scholarly principles, practice and procedures. This framing is useful for illustrating how the methodology serves the objectives of this research project

#### **4.2.2 Principles**

All design science research projects must operate under a set of common principles (Rowe, 1991). The key distinguishing characteristic of design science, as opposed to other research methodologies, is that new knowledge and practical solutions are produced through the design of some ‘artefact’. Design science artefacts may be constructs, models, methods or instantiations (Williamson & Johanson, 2013). The current research focuses on the design of a measurement model. The design of the model follows the principle of iteration, which refers to alternating between phases of design, testing, and evaluation. Plomp and Nienke (2007) describe this principle as the successive approximation of theory and practical products. It has also been described as prototyping (Baskerville, Pries-Heje, & Venable, 2009). The purpose of the iterative design process is to produce an artefact with both theoretical and practical value. Each redesign of the model is evaluated against theoretical and practical criteria, as well as the study’s research questions. Value creation is further aided by another tenant of the methodology—collaboration. From the outset of the project, design science should include the cooperation of researchers, practitioners and intended beneficiaries (Reeves, 2006).

Another tenant of design science is the commitment to generate design principles for the artefact produced. Plomp and Nienke (2007) explain that design research processes should shape the formation of design principles that describe key characteristics of the artefact, explicate guidelines for re-design or replication, and provide theoretical justifications for the artefact and its components. The authors explain that design principles also serve as evidence for the ability of the artefact to provide an effective solution to the research problems.

Shavelson, Phillips, Towne, & Feuer (2003) explain that design science research must also share common principles with other forms of rigorous research. The authors explain that design researchers, like all researchers, must operate under principles and norms of scientific research. Such principles include:

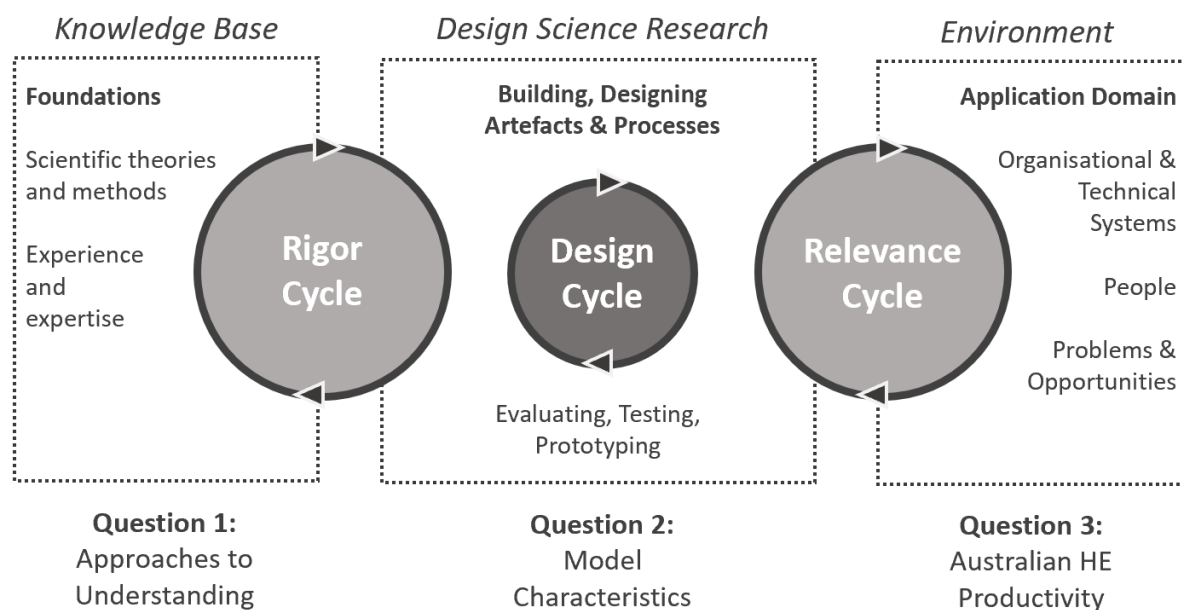
- Posing significant questions that can be investigated
- Linking research to relevant theory
- Using methods that permit direct investigation of the question
- Providing a coherent and explicit chain of reasoning
- Replicating and generalize across studies
- Disclosing research to encourage professional scrutiny and critique

### 4.2.3 Practice

The practice of design science may be understood in terms of three corresponding and interlinked cycles. Hevner (2007) describes the cycles as the principle governing framework behind design science methods. The three concurrent cycles include: (1) the relevance cycle, (2) the rigor cycle, and (3) the design cycle. “The relevance cycle bridges the contextual environment of the research project with the design science activities. The rigor cycle connects the design science activities with the knowledge base of scientific foundations, experience, and expertise that informs the research project. The central design cycle iterates between the core activities of building and evaluating the design artefacts and processes of the research” (Hevner, 2007, p. 88).

Figure 9 illustrates the cycles and their correspondence to the research questions posed in the previous chapter. The research questions Q1, Q2 and Q3 target theoretical, technical and contextual problems related to university productivity, respectively. The knowledge base represented in the figure corresponds to prior scholarly work on productivity measurement and provides initial directions for approaches to understanding higher education productivity. The representation of design science research targets the development of appropriate model specifications for Australian universities via iterations of design and evaluation. The representation of environment corresponds to characteristics and stakeholders of the Australian higher education system, and who participate in developing the model to measure performance within the system.

**Figure 9: Design Science Research Cycles (Hevner, 2007)**



### 4.2.4 Procedures

This section sets parameters around the prototyping phase of research. Different types of artefacts require different protocol and procedures for testing and evaluation. The artefact in this research is a model for measuring productivity change of universities. Williamson and Johanson (2013) explain that models are defined as conceptual objects composed of constructs and construct associations. The constructs and their associations must each

represent real-world phenomena and their respective interrelations. The authors thus specify four criteria for evaluating models during the iterative testing and evaluation phases of design. The criteria listed below ensure a design focus on relevance and utility for the application environment, technical fidelity in design, and scholarly contribution to the knowledge base. The evaluation criteria are thus specified as the following:

1. model constructs assessment
2. construct associations assessment
3. model boundaries assessment, and
4. novelty, revelation and importance assessment

Evaluating constructs involves checking for rigorous definitions of constructs and identifying relationships between the constructs and their real-world instances. Evaluating construct associations involves checking rigorous definitions of the associations and the nature and appropriateness of the associations. Evaluating model boundaries involves determining whether boundaries are clear. It involves determining what the model represents, does not represent, and the model's scope for insight. Evaluating novelty, revelation, and importance involves determining what the model illuminates that was not previously known. If phenomena were already well known, the evaluation determines whether the model provides deeper insight into issues or concepts that were not previously clear or well-understood (Williamson & Johanson, 2013).

## **4.3 Research Design and Methods**

### **4.3.1 Design and Framework**

The research project design incorporates the design science principles and guidelines discussed above. The project design is organised into a framework comprising eight stages. The stages embody all activities required to address this study's research questions. A visualisation of the research framework is shown in Figure 10 with concise descriptions of each stage. The first stage encompasses the background research required to form the project's research questions. The second stage begins the prototyping work with the measurement model. Subsequent stages involve testing, evaluating and improving the productivity measurement model. The first model prototype is informed by the literature review. Contextual factors inform the initial model in as much as the accessible time-series data used to measure productivity change. Beginning in stage four, higher education stakeholders and experts are included in the study to help evaluate the model. Experts and stakeholders are interviewed, and their responses generate insight for further model development with respect to the Australian higher education context. Stages seven and eight offer principle findings and contributions to knowledge and practice based on model measurement results and interview responses.

### 4.3.2 Methods

The following paragraphs summarise the specific methods employed during each stage. Complete details about methods follow in subsequent chapters. The description of methods begins with Stage 2, the initial measurement model design following the literature review and research question formation. Subsequent to the following paragraphs, Table X illustrates how the research stages correspond to the remaining chapters in this dissertation.

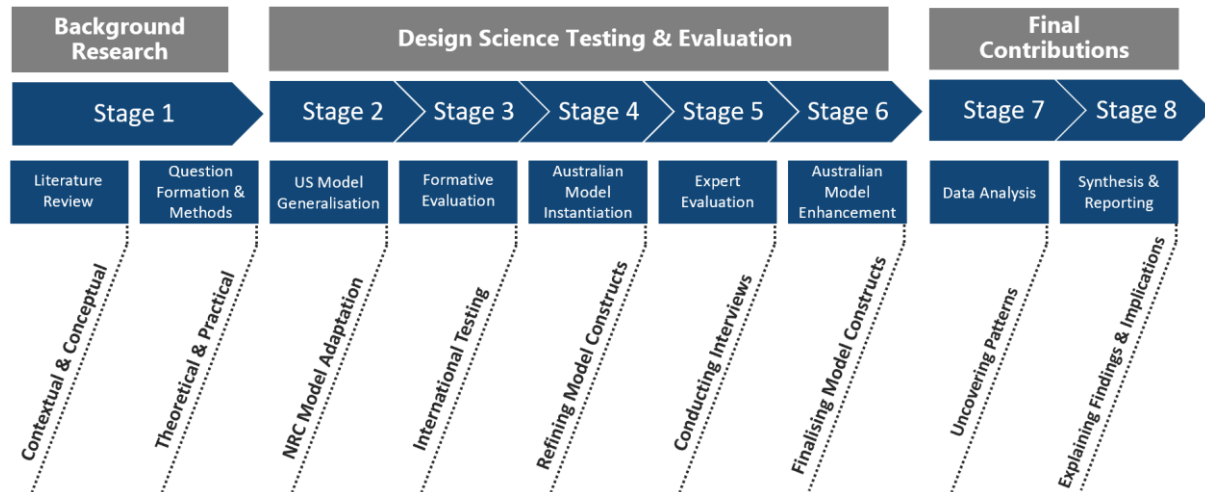


**Figure 10: Eight-stage research project framework**

Stage 2 serves to generalise a productivity measurement model developed for use in the United States. The initial reference model was developed by the United States' National Research Council (NRC) to measure total factor productivity (TFP) change using time-series data. The NRC model was designed to be used on university data from the United States. Specifications are discussed in detail in the next chapter, but it requires adaptation to be used in broader international contexts. The NRC model was chosen because it shows unique potential to generate new insight about productivity in higher education with respect to filling specific gaps in knowledge identified in Chapter 3.

The objective of the initial model design and adjustment of NRC specifications is to first generalise the model for broad international use. This research stage is possible because of an opportunity with the Asia Productivity Organisation (APO), which needed a measurement model to estimate productivity in nine countries in Asia-Pacific. Details about this collaboration are provided in Chapter 5. The rationale for the collaboration is that, prior to this research project, no adaptation or international use of the NRC model had been documented. Thus, making initial judgments about how the NRC model should be adapted to suit the specific Australian context are not straight forward. Before adapting the NRC model for use only in the Australian-specific context, it is useful to generate broad insights about the model's inherent strengths and limitations, and these may be exposed if implemented in a variety of contextual circumstances, each with unique interests in performance and productivity. The model was accordingly generalised for use on the data available in nine APO participating countries and Australian.

Stage 3 is the formative evaluation of the generalised NRC model. International practitioners from nine APO countries use the model on data they had collected, and the model is tested on Australian university performance data collected and aggregated from multiple public

sources. Evaluation of the model is then performed using the four design science model assessment criteria mentioned in Section 4.2.2. Productivity measurement reports provided by the APO practitioners and results from implementing the model on Australian data constitute the evidence used in the evaluation. Evaluation results are used to judge how to continue with prototype redesign and to begin tailoring the model for Australian universities.

Stage 4 is the Australian model instantiation. This stage focuses on the results generated in the previous stage. Broad learnings about the NRC model's general strengths and limitations were exposed with respect to its use outside the US context. Several options for model redesign are entertained based on findings from the evaluation. Improvements are made to both model constructs and construct associations. Adjustments are made to better align with the Australian-specific higher education context and data availability. Further performance data on Australian universities is also collected to improve accuracy and precision in model results.

Stage 5 is the Australian model expert evaluation. This stage is one of the most substantive stages of research in the project. Key informant interviews are conducted with 20 Australian experts and stakeholders to provide feedback on the updated Australian model and its results, as well as to document interests and concerns about University productivity and performance in the country. The interview participants represent a purposive sample of experts and stakeholders, and interviews were recorded and transcribed. Details about the interviews and discussion guide are provided in Chapter 6. Using participants' responses as new evidence, the four design science model assessment criteria are used again to perform a second comprehensive evaluation of the measurement model. The new technical and contextual information on productivity measurement provided by interviewees exposed further strengths and limitations of the model and also informed how measurement results should be interpreted in context.

Stage 6 concludes the prototyping phase of research. It involves making final enhancements to the measurement model based on the evaluation of the model informed by Australia experts and stakeholders. Insights from the evaluation are used to adjust both model constructs and construct associations to increase its fit-for-purpose in the Australian context. The result of this stage of research is the final version of the measurement model in this research to be used on Australian higher education data. The model is intended to address diverse stakeholder needs and perspectives and to estimate university productivity with more validity and reliability than predecessor models in this research or prior models in the scholarly literature. This is possible because of the final version's flexibility in representing different specifications of productivity, as outlined in the measurement model taxonomy in Section 3.4 of Chapter 3. These alternative specifications include SFP and TFP, as well as joint and separate productivity. Details about final measurement model specifications and additional collection of university performance data are provided in the final section of Chapter 6.

Stage 7 is the comprehensive analysis of measured results. Results are available after the final version of the model is run on Australian university data. Nuanced trends and patterns are identified among institutions within the Australian system. The final model is comprehensive in the information it considers, and results are easily decomposed for deeper insight. This allows for an array of analyses to be conducted, the results of which provide the basis for a comprehensive assessment of Australian university productivity. The assessment exposes

emergent patterns among institutions with respect to what is driving productivity growth and change. A threshold analysis is also conducted on the model. The threshold analysis explores the range of plausible interpretations of the reported quantitative data and provides great value for explaining trends and revealing limitations in the data. The analysis continues to explore alternative interpretations of the data based on adjustments made to model parameters. The suite of results inform more robust quantitative findings, and when paired with qualitative evidence, authentic patterns and behaviour within the system can be identified.

Stage 8 is the final stage and involves synthesising and reporting all findings, as well as addressing the research questions. Findings from the final quantitative results and each subsequent model evaluation are examined to define the design principles of the measurement model and the aspects of the model that represent those principles. Evidence from each productivity analysis mentioned above is synthesised to make conclusions and inferences about productivity trends across the higher education system. Implications of all findings are delineated in terms of policy, practice, and technique. The chapter concludes with recommendations for continued and future research.

## **4.4 Summary**

### **4.4.1 Focus and Contributions**

This chapter explained the research methods and methodology of this project. It discussed how the remainder of this dissertation serves to document an iterative design and evaluation cycle of a new higher education productivity measurement model. The final model uses indexing techniques and thus produces estimates of Australian university productivity change. The quantitative findings from the measurement model, along with responses from 20 key-informant interviews, compose a body of evidence used to inform assessment of Australian higher education productivity and to make conclusions about higher education productivity measurement and assessment in general.

The chapter provided a rationale for how the methods address this study's research questions. The chapter stressed that Question Q2 carries a central investigative focus throughout the project. The second question regards appropriate characteristics for an Australian higher education productivity measurement model. The focus on this question is what justifies methods for prototyping a new measurement model. This does not imply, however, a disproportionate emphasis on the value of Question Q2 in relation to the others. Rather, the chosen methods are instrumental for addressing all three questions in a complete manner. Addressing the questions is achieved with the implementation of design science research practices. Figure 9: Design Science Research Cycles (Hevner, 2007) from this chapter captures the design science methodology by illustrating three concurrent cycles of inquiry. The cycles are the knowledge, design and relevance cycles, which directly correspond to the three research questions.

The research can be organised into eight stages—each of which follow design science standards for research practice and procedure. Stage 1 incorporates the literature review and research question formation. Stage 2 through Stage 6 represents the measurement model prototyping phase of research, which involves successive versions of the model—designed, tested and evaluated using the methods described above. Stage 7 involves the analysis of all data and results provided by the final version of the measurement model. Stage 8 reconciles all qualitative and quantitative findings to frame authentic representations of Australian university productivity. The final stage culminates in key contributions to knowledge and theory, as well as implications for policy and practice.

#### **4.4.2 Considerations and Limitations Arising**

As identified in the literature review, the use of indexing methods represents a promising opportunity for advancing knowledge and practice in higher education productivity measurement. Indexing methods, however, are best suited for measuring changes in productivity, not absolute productivity. The implication is that measurement results can only be used for assessing relative rates of institutional productivity growth or decline, rather than relative levels of overall performance. This places a limitation on the specific questions that can be addressed with the measures produced. Further, the purposive sample of key informants who participate in the interviews does not represent a random sample from a specified population. Participants were selected based on professional expertise, decision-making responsibility, and stake in productivity measurement results. Results from the analysis of interview response data should therefore not be mistaken as definitive conclusions about the phenomena they represent. Rather, they represent a critical mass of informed opinions and perspectives on the questions raised in the interview discussion guide. Results constitute purposive reflections on the model, its results and broad interests in Australian university productivity. The range and extent of the issues discussed reflects the point at which data saturation occurred with responses. The relative importance of issues raised, however, cannot be determined based on the frequency of responses, as the sample was not representative.

**PART II:**

**MODEL DESIGN, TESTING & EVALUATION**

## CHAPTER 5: INITIAL MEASUREMENT MODEL DESIGN

*On the basis of its structure, go with the Törnqvist Index... the more structure you build into a model, the better the results.*

–Study Participant TE.9  
Former University Chief Financial Officer

### **Abstract:**

The contribution of this chapter is the establishment of an initial framework for the prototype measurement model in this research and to establish a base of knowledge for addressing research question Q2. The chapter gives a rationale for the framework chosen. Model characteristics are specified (A) to align with insights from Chapter 3 and (B) to ensure fit-for-purpose with respect to the first application and test of the model. The first test is facilitated by the Asia Productivity Organisation (APO) on data from nine Asian countries, as well as on Australian data. The chapter explores technical literature on measuring productivity change with indexing methods to explain how data is treated in the model. A model proposed by the United States National Research Council (NRC) is chosen as a baseline. Adaptations to the NRC model are also presented to illustrate how measurement practice is designed to align with the needs of APO higher education practitioners and data accessible in Australia. Notable insights generated in this chapter may also be found in Moore et al. (2018b).

### **Key Words:**

Törnqvist index, productivity change, inputs, outputs, joint productivity

## 5.1 Introduction

This chapter presents developments and insights from Stage 2 of the research. In the context of the overall project, the contribution of this chapter is to establish a foundation of knowledge for addressing research question Q2, concerning productivity measurement model specifications appropriate for Australian university contexts. Key contents of this chapter are mathematical in nature. Technical details are presented in terms of model constructs and construct associations to allow for targeted evaluation in subsequent stages of research (Williamson & Johanson, 2013).

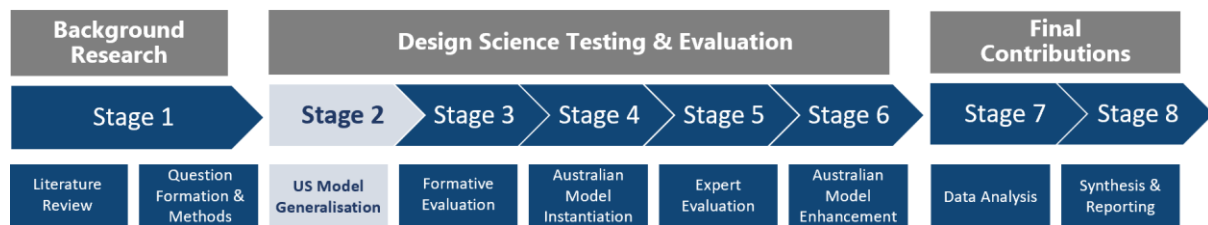


**Figure 11: Research Stage 2**

The chapter proceeds to explain how initial modelling decisions were made. Key decisions were based on the measurement contexts and interests of Asia Productivity Organisation (APO) member countries whose practitioners test the first version of the model. Decisions also hinge on the technical opportunities for research identified in Chapter 3. Five model dimensions were explained, and gaps in knowledge were identified based on those dimensions: (1) time differential, (2) input characteristic, (3) output characteristic, (4) production complexity, and (5) data aggregation technique. Table 6 summarises model specifications that address those gaps in knowledge. The first prototype measures ‘productivity change’. It measures total factor productivity (TFP) using university financial variables as inputs. Multiple research and education outputs are considered. By default the model measures joint production of education and research, but some countries adapted the method to approximate research and education productivity separately. The data aggregation method relies on indexing techniques.

**Table 6: Initial prototype measurement model specifications**

| Modelling Dimension     | Specification                                            |
|-------------------------|----------------------------------------------------------|
| Time Differential       | Productivity change                                      |
| Input Characteristic    | TFP                                                      |
| Output Characteristic   | Multiple outputs                                         |
| Production Complexity   | Joint productivity (separate productivity approximation) |
| Data Aggregation Method | Törnqvist Index                                          |

## **5.2 Initial Design Decisions**

### **5.2.1 Measurement Opportunities**

Chapter 3 revealed opportunities to develop higher education productivity measurement. Indexing methods were identified as prime aspect of modelling to investigate further. Key reasons include (1) their flexibility in estimating productivity under circumstances when limited information is available about production technology, (2) few required assumptions or limitations needing to be imposed on data, and (3) simple mathematical techniques allowing for decomposition of results and the ability to explore relationship dynamics between data elements. Until recently however, there were very few options for aggregating higher education's multiple, unpriced outputs into a reliable composite output indicator (Carrington et al., 2005). A report by the United States National Research Council (NRC), however, released a comprehensive report advancing the Törnqvist Index (TI) as a preferred indicator for measuring higher education productivity change and provided methods for aggregating at least two different types of higher education outputs (Sullivan et al., 2012). The advancement provides clear directions forward for continued research.

The NRC guidelines define a composite education output indicator of 'adjusted credit hours', which incorporates student course load and graduate completions. The indicator is based on conceptual and empirical foundations. The incorporation of completions with student course load captures a 'sheepskin effect' that represents additional value gained when students earn a degree. Calculation of the adjusted credit hour indicator is based on findings by Jaeger and Page (1996) and Park (1999). The two studies each generate estimates of the value of university degree attainment based on both human capital and signalling theories of education. Further, the NRC incorporated measurement standards established by the OECD (2001) and the US Bureau of Labor Statistics (BLS) to specify the aggregation of multiple inputs into a conventional TFP indicator. Broadly, the standards specify inputs of labour expenditure, capital expenditure, and intermediate expenditure. Details are provided in Section 5.5.

A prime contribution of this chapter—also explicated in Moore et al. (2018b)—is a further adaptation to the NRC model, which allows for the concurrent aggregation of research outputs with education outputs. Limitations for aggregating multiple outputs using indexing methods raised in the literature provided the first impetus to continue work on the NRC model. Contextual reasons discussed below, however, provided a clear and urgent rationale for developing more comprehensive data aggregation methods based on the TI. As described in Chapter 3, TI methods calculate the change in a composite output index and divide it by a composite input index. Usually these indexes are calculated from priced inputs and outputs. The prime contribution of this chapter is to provide a method for aggregating non-priced outputs. Details are given in section

### **5.2.2 Contextual Opportunities**

Initial design decisions were made to ensure that the first iteration of the measurement model was fit for purpose across nine Asia-Pacific countries in an international field test performed by APO higher education practitioners. The countries involved are Cambodia, Fiji, India,

Indonesia, Malaysia, Pakistan, the Philippines, Sri Lanka, and Thailand. One of the supervisors of this dissertation, Professor Hamish Coates, was named Chief Expert to oversee the initiative. Coates (2017a) describes early work in the project, where discussion uncovered the possibility of using indexing methods inspired by the NRC. The literature review for this dissertation, which followed early project coordination, then demonstrated the existence of mutual interests in first designing an international version of the NRC model for the APO initiative to later be developed further for use in the Australia specific context.

The APO partnership marked a significant opportunity to include stakeholders and practitioners in the design process of the model being developed in this dissertation. The APO sponsored and supported the effort for its member countries to measure the productivity of their respective higher education sectors using the adapted NRC model. Practitioners from each country were tasked with providing contextual overviews of their respective higher education systems, trends and needs. The principle researcher of this dissertation was tasked with defining the technical specifications of the measurement model. Full details of the APO initiative are available in (Coates, 2017a), but initial consultations with country practitioners revealed three overarching adaptations to the NRC model that would need to be made in common to each participating country. Without the three adaptations the model would not have enough flexibility to be used for purposes and contexts outside the US. The adaptations are listed below, explained in the paragraphs that follow, and technical specifications are provided in sections 5.4 and 5.5.

1. Education output is calculated based upon full-time student load instead of US ‘credit hours’
2. Research output indicators are introduced
3. Production complexity is changed from separate education output production to joint production of both research and education outputs

First, the base model can be internationalised by designating the primary education output as the total number of full-time or full-time equivalent students (FTEs) in a given year, instead of United States ‘credit hours’. United States universities associate each subject delivered with an assigned number of credit hours or ‘points’ which are loosely determined by how much time students should spend attending lectures for that subject during a given week. Students’ full- or part-time status in US universities is determined by the number of credit hours they take. Thus, an institution’s total yearly student load, or total number of yearly FTEs, indicates the scale of educational operations. Internationally, although subject point calculation methodologies vary, most institutions track their yearly student FTEs, so this indicator allows enough consistency and accuracy across countries to represent total output of education delivery during a given year. The NAS model calculates a number of ‘adjusted credit hours’ for its final function output for education, which also accounts for the total number of graduate completions by coursework during the same year. The model adaptation using FTEs also accounts for graduate completions by coursework and calculates ‘adjusted FTEs’ in the exact same manner as Sullivan et al. (2012).

Second, APO practitioners expressed the desire to track research productivity. Coates (2017a) arrived at potential research indicators to serve as proxies for outputs based on data collected by the counties. The indicators include publications, citations, patents, research completions,

and research funds. The adapted model also includes new protocol for how to aggregate these different types of data elements into a composite output indicator.

Third, the incorporation of research output requires further generalization of the input-side of the NAS model. The financial inputs for the NAS model rely on a unique university accounting initiative in the United States in which institutional costs are tracked with respect to academic function, such as education, research, or administration (Sullivan, 2012: 68). Thus, the financial inputs to the NAS model include only direct costs allocated to education activity. Most countries' higher education systems and institutions, however, do not employ accounting methodologies for tracking costs by academic function. By incorporating both education and research output indicators, the adapted model allows for inclusion of financial inputs that need not be separated by academic function. The primary input categories for the current study's model are the same as the NAS model—labour, capital, and intermediate—but costs associated with these inputs are not divided by academic function. The current model assumes all inputs lead to joint production of education and research outputs.

### 5.2.3 Model Usage Guidelines

After country consultations, all participating APO countries were encouraged to use the same input-output framework, and not to add new data element types to the framework. The input-output framework consists of 17 possible inputs and outputs to use in the calculation. Guidelines given to practitioners for calculating their productivity change indexes intended to simplify the calculation, but participants were aware of the complexities involved in aggregating the data. They had some flexibility to adapt the model to local circumstances. Details of the full framework and aggregation method can be seen in **Appendix A**, as well as in Coates (2017a) and Moore et al. (2018b). It may be useful to read section 5.4 before Appendix A. Details concerning country-by-country data collection and analysis of results are provided in the next chapter with respect to evaluation of the first iteration of the model.

For subsequent discussion, the next section explicates how the adapted NRC model was applied to Australian higher education data. Some of the data elements included in the APO practitioner input-output framework, such as 'learning outcomes', 'graduate employment', and 'patents' were either inaccessible or were deemed inappropriate for an endogenous assessment of institutional productivity in Australia. Thus, for the sake of consistency over the course of tracking Australian model design iterations, only the data elements and calculation techniques used on Australian data are discussed in depth.

## 5.3 Model Constructs

### 5.3.1 Constructs Overview

It is helpful to first reiterate the language used by Williamson & Johanson (2013) from the previous chapter used to describe the basic components of any model. Models comprise *constructs* and *construct associations*. This language will henceforth be used to refer to the

specifications of the productivity measurement model under development in this research. The purpose is to frame all model specifications in terms of the criteria by which it will be assessed in the next chapter. Corresponding to Table 6, the prime model specifications relating to the taxonomy from Chapter 3 may be thought of as the model's overarching constructs, or the prime building blocks of the model. That is, the adapted NRC model measures productivity change using TFP input aggregation and multiple outputs. It measures education and research productivity jointly using indexing methods. These main constructs will be the subject of assessment over multiple evaluation iterations. The bulk of assessment however, will focus on the constructs that compose the model in practice—that is the input and output indicators that make constitute the model's input-output framework. Further, assessment will focus on the associations between these constructs—that is, the mathematical correspondences between data elements as defined through TI methods. The following sections provide complete details for all input and output constructs use in the model, as well as the associations between them.

### 5.3.1 Input Constructs

**Table 7: Input constructs and specifications<sup>1</sup>**

| Construct Type | Construct              | Variable       | Data element                           |
|----------------|------------------------|----------------|----------------------------------------|
| Labour         | Academic staff         | L <sub>1</sub> | Academic staff salary and benefits     |
|                | Non-academic staff     | L <sub>2</sub> | Non-academic staff salary and benefits |
| Capital        | Land                   | K <sub>1</sub> | Land capital services                  |
|                | Buildings              | K <sub>2</sub> | Buildings capital services             |
|                | Equipment & other      | K <sub>3</sub> | Equipment and other capital services   |
|                | Repairs & maintenance  | K <sub>4</sub> | Repairs and maintenance                |
| Intermediaries | Administration & other | I <sub>1</sub> | Administration and other expenses      |
|                | Grants & scholarships  | I <sub>2</sub> | Grant and scholarship expenses         |

The model input constructs (Table 7) are based on those of the NRC model. Sullivan et al. (2012). The prime input constructs are labour, capital and intermediaries, and the variables L, K and I refer to them respectively. Each construct is represented by data elements from university annual financial reports. Annual data on university assets and expenditures signify the magnitude of resources used and consumed by institutions to produce outputs over the course of a calendar year. Although specific accounting practices may differ across international contexts, labour, capital, and intermediate input constructs are assumed to be universal to any physical higher education institution. There are of course limitations in considering only financial indicators as inputs. The student, for example, arrives in an institution with some prior knowledge and capacities that are enhanced by the institution. In this sense, the student and institution collaborate to produce the final product. Such dynamics are not captured by the current model, but instead provide an overall picture of productivity (Sullivan et al., 2012).

---

<sup>1</sup> Data is obtained from the Australia Government's publications on university finances (Department of Education and Training (DET), 2018b).

Labour expenditures include annual wages and benefits for academic and non-academic staff members. Intermediate expenditures include purchased goods and services, such as energy and materials, as well as annual administrative expenditures in addition to direct labour costs. Those listed in the framework above correspond to figures available in Australian institution financial reports. Capital expenditures are derived from both university balance sheets and income statements. They include items such as buildings, equipment and rates of depreciation.

Unlike labour and intermediate expenses, fixed capital assets do not represent annual expenditures. They do, however, contribute to annual output production. Hence, an additional construct is needed to account for the value that capital assets provide to production process on an annual basis. The construct to represent annual capital expenditures is ‘capital services’, also referred to as ‘rental value of capital’ or ‘capital flows’. Estimating capital services for productivity change measurement is common practice (ABS, 2012; BLS, 2007; OECD, 2001). Sullivan et al. (2012) explain that the rental value of capital should cover both depreciation and real rates of return of comparable investments. The OECD (2001) provides multiple methods for estimating capital services. Each method revolves around estimating a fraction to be multiplied against the full value of capital assets. The purpose is to estimate the effects of changes in fixed capital assets on production processes being measured on yearly time-spans. Total value of capital assets is often much greater than annual expenditures, and the full value of capital assets does not represent accurate annual contribution to year-on-year production processes. The estimated fraction may be referred to as the ‘capital service factor’.

In the absence of reliable data in Australia on rates of depreciation and real rates of return in the education sector on fixed capital, the model uses a simplification. It generalises from the NRC model and uses a baseline assumption that the average value extracted from fixed capital on a yearly basis is five per cent of the total value of the assets. The value of five per cent serves as a placeholder in the first iteration of the model to allow for more valid estimates yearly changes in capital flows than using the full book value of capital. The capital service factor is acknowledged as a model construct to be improved upon for later iterations of the model.

### 5.3.2 Output Constructs

**Table 8: Output constructs and specifications<sup>2</sup>**

| Construct Type | Construct              | Variable       | Data element                                       |
|----------------|------------------------|----------------|----------------------------------------------------|
| Education      | Student Load           | E <sub>1</sub> | Number of equivalent full-time coursework students |
|                | Coursework completions | E <sub>2</sub> | Number of coursework graduates                     |
| Research       | Publications           | R <sub>1</sub> | Number of weighted publications                    |
|                | Research completions   | R <sub>2</sub> | Number of research graduates                       |
|                | Research income        | R <sub>3</sub> | Amount of research income                          |

The prime model output constructs are education and research, and the variables E and R refer to them, respectively. Inclusion of research outputs represents the prime deviation from the NRC model in university productivity measurement. The education and research output constructs in the adapted model are relevant for both the Australian context and the APO country contexts. The paragraphs that follow discuss all output constructs listed in Table 8.

For education the model uses equivalent full-time student load (EFTSL) and graduate completions to serve as the primary education output constructs. The EFTSL figure is a standard unit of measure for the Australian system. The inclusion of EFTSL represents an adaptation to the NRC model, which uses ‘credit hours’ instead of load. Credit hours in the United States system are assigned to individual subjects. A single subject taken over the course of one semester, such as Calculus I, would be worth typically three or four credit hours. The credit hour is a measurable output, and the number of credit hours a student takes in a year determines his or her full-time or part time status. When all credit hours delivered by an institution are aggregated, the scale of educational service delivery can be inferred. In Australia, however, all universities do not follow the unifying system of credit hours for subjects delivered. It is a more aggregate measure than credit hours, but the total EFTSL count for an institution also serves as a proxy for the scale of educational service delivery during the academic year.

Just as with the financial input variables, there are limitations to the output. Total annual EFTSL and graduate completions provide a reliable starting place for measuring education productivity. The full value that course programs and the degree attainment provides to the individual and to society are not captured by these constructs. The indicators have flaws, but they are common, standard units of measure. They cannot provide a complete representation of education outputs, but they do represent a significant part of the story (Sullivan et al., 2012).

The original NRC framework intentionally excludes research outputs. The model measures only education productivity. Measuring separate education productivity is possible because accounting practices in the United States allow for corresponding estimates of educational

---

<sup>2</sup> Education outputs are obtained from the Australian Government’s database on higher education statistics (DET, 2018c). Research outputs are obtained from the database on time-series research data (Department of Education and Training (DET), 2018a).

inputs. Total amounts of expenditure can be disaggregated for allocations to different academic functions. This allows for direct association of educational inputs and educational outputs. Accounting practices in Australia, however, do not allow for separating research and education costs. No standard practice for doing so exists across all Australian institutions. The alternative approach is to link all outputs to all inputs using a joint productivity model.

There are many limitations with research output indicators, but the three included reflect the key performance indicators used by the Department of Education and Training (DET) to award research funding to institutions during the 2007 to 2013 period analysed in this initial version of the model (DET, 2017a). Further, examining research income and graduate research completions along with publication data speaks to a broader set of interests in research. Gralka, Wohlrabe, and Bornmann (2019) explain that governments are interested in seeing research impact with respect to other sectors of society in addition to its importance to academia. Thus, for the Australian context, the chosen research outputs represent data collected, reported and used by universities and the DET for key decision-making and broader goals.

Worth noting is that research income could be thought of as an input for the conduct of research, but in the Australian case, it represents the outcome of competitive processes to generate financial support for subsequent research activity. As with the education outputs, the research outputs as well cannot encompass the total value of research for individuals, institutions, and society. No input-output combination for universities could portray productivity as definitively as can be done for firms in the private sector, which have priced outputs and objectives to maintain or maximize margin. Significant information, however, can be derived from the input-output analysis allowed by this framework, even if incomplete.

## **5.4 Construct Associations**

### **5.4.1 Construct Associations Overview**

This section is mathematical in nature. It defines Törnqvist indexing, and it explicates all technical adaptations made to the NRC model. At the end of this section, a visualisation is provided to illustrate how the TI serves to link and aggregate all data elements in the model to generate an aggregate productivity change index. The figure is intended as a supplement to explain all construct associations without needing precise mathematical definitions. For quick interpretation of the figure, please refer to the ‘variable’ columns in input and output tables above, which refer to the unique data elements used in the model.

### **5.4.2 Input Associations**

The TI calculation technique defines the associations between input and output constructs in the model. A TI productivity change index involves calculating input and output indexes separately and then dividing the output index by the input index. The section first explicates the input index. In this model, all input variables are expenditures. The index  $X_t$  for costed

inputs from time  $t - 1$  to  $t$  can be seen below. The notation differs from the summative, log-linear form shown in (Moore et al., 2018b). For simplicity, it is represented below in multiplicative, multinomial form.

$$X_{\Delta t} = \prod_{i=1}^n \left( \frac{x_{i,t}}{x_{i,t-1}} \right)^{u_i}, \quad [5]$$

where  $n$  is the number of input types an institution uses,  $x_{i,t}$  is the total cost of input  $i$  at time  $t$ , and  $x_{i,t-1}$  is the total cost of input  $i$  at time  $t - 1$ . The term  $u_i$  serves as a value weight for each type of input. It is the arithmetic average of the value share of each input type, as a proportion of total input costs over the two time periods. The formula is shown below:

$$u_i = \frac{S_{x_{i,t}} + S_{x_{i,t-1}}}{2}, \quad [6]$$

where  $S_{x_{i,t}} = x_{i,t} / \text{total input costs } (t)$ ,  
and  $S_{x_{i,t-1}} = x_{i,t-1} / \text{total inputs costs } (t-1)$ .

The variable  $S$  refers to the ‘share’ of value for each construct at the designated time.

In the case of the current model,  $n = 8$ , but data elements are grouped by construct type to allow for Törnqvist ‘chain indexes’ (BLS, 2007). The chain indexing method in this study requires first calculating component indexes for labour, capital, and intermediaries separately. Then a single composite input index may be calculated from the component indices. Consider the example of the labour component index. is as follows. Let  $L_t$  signify the TI component index for labour expenses from time  $t - 1$  to time  $t$ .

$$L_{\Delta t} = \left( \frac{L_{1,t}}{L_{1,t-1}} \right)^{u_{L_1}} * \left( \frac{L_{2,t}}{L_{2,t-1}} \right)^{u_{L_2}} \quad [7]$$

The component indexes for capital  $K_t$  and intermediaries  $I_t$  can be calculated in the same way. Using a chained TI process, the composite input index  $X_t$  may be calculated using each component input index. The formula is shown below.

$$X_{\Delta t} = (L_t)^{u_{L_t}} * (K_t)^{u_{K_t}} * (I_t)^{u_{I_t}} \quad [8]$$

The  $u_{j_t}$  value weights for the component indexes are calculated using the sum cost of each construct component within the respective  $j$  construct type, labour, capital, or intermediaries. The example for  $u_{L_t}$  is shown below.

$$u_{L_t} = \frac{S_{L,t} + S_{L,t-1}}{2}, \quad [9]$$

where  $S_{L,t} = (L_{1,t} + L_{2,t}) / \text{total input costs } (t)$ ,  
and  $S_{L,t-1} = (L_{1,t-1} + L_{2,t-1}) / \text{total inputs costs } (t-1)$ .

### 5.4.3 Output Associations

A prime contribution of this chapter responds to concerns raised by Carrington et al. (2005) about the limitations of indexing methods to aggregate non-priced outputs. The initial model for this research proposes the following solution. a non-priced output change index from time  $t - 1$  to time  $t$  may be given by:

$$Y_{\Delta t} = \prod_{i=1}^m \left( \frac{y_{i,t}}{y_{i,t-1}} \right)^{v_i}, \quad [10]$$

where  $m$  is the number of specified output data elements,  $y_{i,t}$  is the amount of output  $i$  at time  $t$ , and  $y_{i,t-1}$  is the amount of output  $i$  at time  $t - 1$ . The term  $v_i$  denotes a value weight for output  $i$ , signifying a proportion of the total value of all outputs. Designating value weights for the output change index differs from the procedure for the input change index. Each  $v_i$  must be designated by the researcher with the constraint that  $\sum_{i=1}^m v_i = 1$ .

The non-empirical nature of the value weights assigned to the outputs is the limitation to which Carrington et al. (2005) refer. In principle, however, value weights could be assigned with respect to strategic importance, stakeholder interests, or on parallel quality and impact evaluations, as discussed by Massy (2012). Carrington et al. (2005) refer to DEA's ability to elicit 'shadow price' information within a set of data using its algorithms to determine weights empirically. Conceptually and practically, however, algorithmic determinations of value weights based on optimisation procedures and the shape of the dataset hold no more validity than researcher or practitioner determinations of value weights based on context or on other exogenous factors not represented in the set of data being fed into the productivity measurement model. The initial version of the model in this research treats education and research outputs as equal. The equal weighting likely does not hold for each institution, but it does have empirical support, as it runs consistent with traditional academic work expectations in Australian universities (Watson, King, Dekeyser, Bare, & Baldock, 2015). It further provides a minimally controversial designation for initial model testing.

With respect to the data elements used in the current mode  $Y_{\Delta t}$  is formed using the composite education and composite research indexes  $E_{\Delta t}$  and  $R_{\Delta t}$ , as shown below:

$$Y_{\Delta t} = (E_{\Delta t})^{v_E} * (R_{\Delta t})^{v_R}, \quad [11]$$

where  $v_E = v_R = 0.5$ , signifying the equal strategic or ideological weighing discussed above. The following paragraphs provide explanations for how further adaptations to the NRC model allow for the calculation of the composite education and research indexes

The primary education output is an association between graduate completions and EFTSL, named 'adjusted load'. The association is an adaptation of the NRC model's 'adjusted credit hours'. Sullivan et al (2012) that the credit hours delivered by an institution represent one of the best measurable indicators for the scale of the education function; however, graduate completions of programs are also a desired deliverable of the institutions, which may reflect the effectiveness or quality of the courses being delivered. Additionally, qualifications hold value in that they can be a signal to potential employers the credibility of the person holding the degree. This phenomenon is handled in the model with a 'sheepskin effect', additional value assigned for degree completion. The sheepskin effect may be different for different fields of study, and different studies have shown different average magnitudes of the sheepskin effect. The magnitude of the sheepskin effect remains an empirical question, and pending further study, the NRC panel recommends that the sheepskin effect be assigned the equivalent to an additional year's worth of full-time study (Sullivan et

al., 2012). This is taken as a baseline for the model developed in this study. The formula for the current model, adapted from the NRC model, is shown below.

Adjusted load = EFTSL + (Sheepskin effect  $\times$  Completions)

Because EFTSL represents the equivalent of one year of full-time study, the baseline for the sheepskin effect in the model is a value of one. Full details and background about these calculations are available in the NRC report (Sullivan et al., 2012). The education component index  $E_{\Delta t}$  from time  $t - 1$  to time  $t$  for the current model is given by the following:

$$E_{\Delta t} = \left( \frac{E_{1,t} + E_{2,t}}{E_{1,t-1} + E_{2,t-1}} \right)^{v_{E_{1,2}}} \quad [12]$$

Where  $E_1$  and  $E_2$  correspond to the data elements in the input-output framework. Calculation of  $R_{\Delta t}$  from time  $t - 1$  to time  $t$  may be given by the following:

$$R_{\Delta t} = \left( \frac{R_{1,t}}{R_{1,t-1}} \right)^{v_{R_1}} * \left( \frac{R_{2,t}}{R_{2,t-1}} \right)^{v_{R_2}} * \left( \frac{R_{3,t}}{R_{3,t-1}} \right)^{v_{R_3}}, \quad [13]$$

where  $v_{R_1} = v_{R_2} = v_{R_3} = \frac{1}{3}$ . Empirical derivation of value weights for these elements might consider DOET's (2017d) methodology for awarding research block grants—as the data elements used in the model's input-output framework are consistent with those used by the Australian federal department of education. However, the scope of the first test of the model is not to arrive at definitively precise productivity growth estimates for the sector.

The final construct association for the model is the relationship between input and output indexes. The final productivity change index from time  $t - 1$  to time  $t$  is given by  $P_{\Delta t} = \frac{Y_{\Delta t}}{X_{\Delta t}}$ .

As stated at the beginning of this section, Figure 12 below provides a visual aid of the model. It shows the multi-step process of Törnqvist indexing to estimate  $P_{\Delta t}$  from the data elements. Each data element in the figure corresponds to those listed in Table 7 and Table 8 above. The figure should be examined from top to bottom. Data elements are first used to create input and output component indexes. Component indexes are then used to create input and output composite indexes. Finally, the composite indexes are used to estimate the productivity change index  $P_{\Delta t}$ .

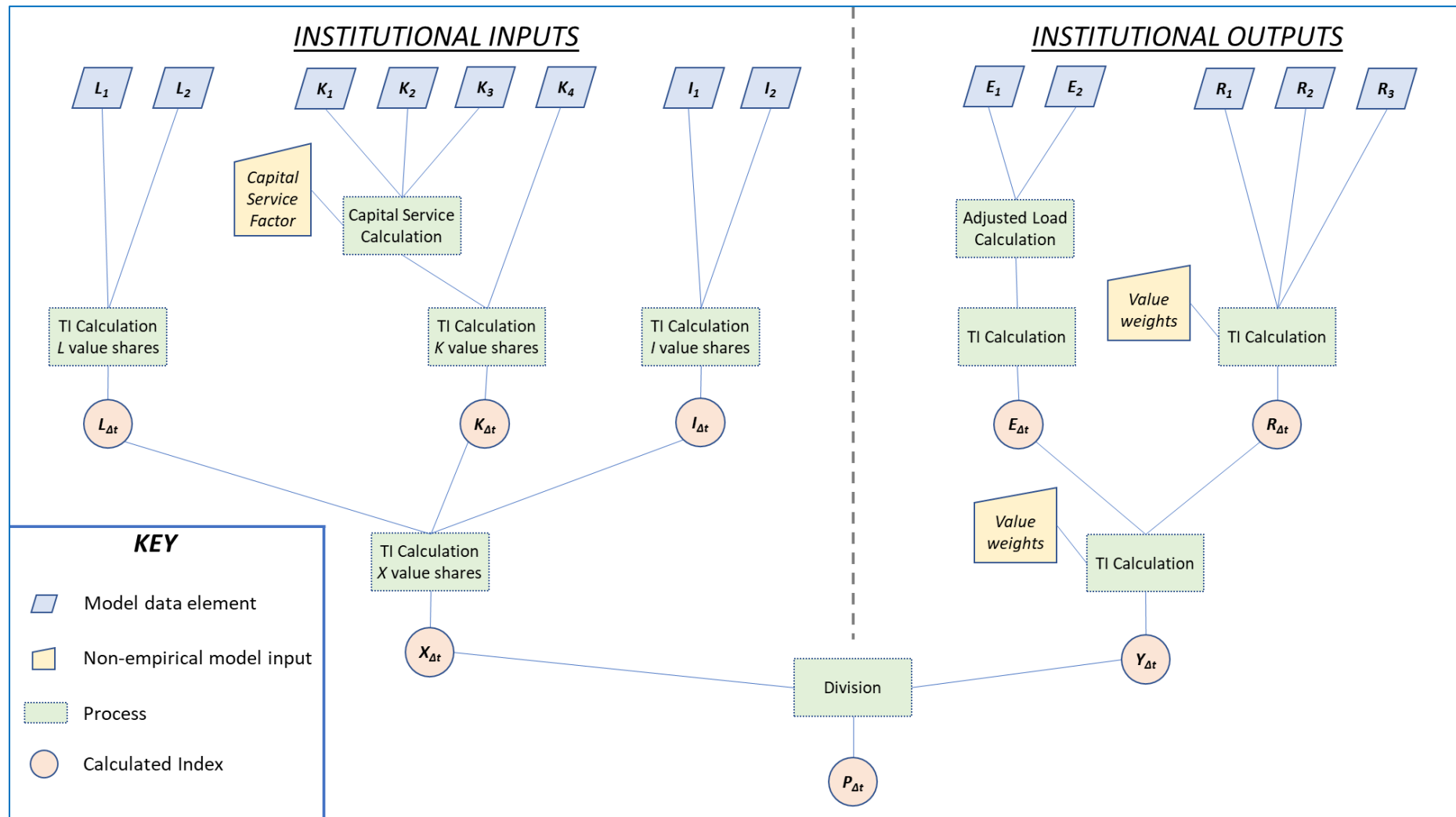


Figure 12: Model construct associations and data aggregation schematic

## **5.5 Summary**

### **5.5.1 Focus and Contributions**

This chapter has made two key contributions to the research in this dissertation. First, it specified a baseline productivity measurement model for further development in this study. The baseline model was first presented by the US National Research Council for measuring productivity change in the US higher education system. Designation of this model includes a reference point for an input-output framework and establishes a measurement technique called the Törnqvist index. Second, the chapter provided a rationale for conducting an initial test of the prototype measurement model on multiple countries' data. A collaboration with the Asia Productivity Organisation connected this research project with practitioners from nine different countries in the Asia-Pacific region. Evaluating the experiences and results of APO practitioners who use the model affords an opportunity to shed light on broad strengths and limitations of the model under a range of measurement contexts and performance interests.

This chapter also specified all model constructs and construct associations of the adapted measurement model to be tested by APO practitioners. Model constructs were presented with respect to the range of data elements included in its input-output framework. Model input and output constructs were chosen based on scholarly literature and on the measurement priorities of stakeholders from participating test countries. Construct associations were presented with respect to the TI methods used to compute productivity change estimates from individual data elements. The benefit of TI methods is their ease and transparency of data aggregation, as well as their precedented widespread use by governments and multi-lateral organisations to measure productivity change across various sectors and industries.

### **5.5.2 Considerations and Limitations Arising**

Benefits of using TI methods for estimating productivity change include their accessibility and transparency. Their lack of advanced mathematical techniques is also associated with limitations. TI methods are not associated with any definitive guidelines for aggregating data elements measured in different units. This means that the researcher retains control over how all data elements are emphasised during calculation. There is, however, no set procedure for ensuring an empirical basis for how value weights are assigned to data elements. When using TI techniques, definitive empirical justifications for value weights on data elements are, in fact, not possible. Judgements can be made, however, in ways that reduce controversy and contention and expand opportunities for insight. This is a matter to be explored throughout the research.

The use of TI methods gives rise to another matter of contention that may affect the fidelity of empirical results. Literature on productivity change indexes warns that, because measured

results are normalised to a value of ‘1’ at the beginning of the period being analysed, change trends can be distorted or misrepresented if institutions being compared are not starting at a reliable performance baseline. For initial measurements and assessments of productivity change, this is not a critical issue, but it does represent a challenge to the precision of results and should be addressed in later versions of the model.

## CHAPTER 6: FORMATIVE EVALUATION AND MODEL REDESIGN

*If you don't understand the maths behind the numbers, then your planning for your institution is inferior.*

–Study Participant KS.5

Former University Chief Operating Officer

### **Abstract:**

This chapter describes the first evaluation of the productivity measurement model and give measurement results on Australian universities following model redesign. The model in this research is first tested on data from ten higher education country contexts across Asia-Pacific and Australia. Learnings from the multi-country test establish a base of evidence for addressing the research questions posed in this project. For Q1, findings highlight country interests and understandings of productivity in context. For Q2, new model specifications are developed. For Q3, preliminary measurements on Australian universities are conducted. Key findings from the chapter include support for proof-of-concept of the model in estimating university productivity change. Specific findings from model assessment inform redesign options including: improved associations between input and output constructs, value weight range accommodation for non-empirical model components, and calculation of cumulative productivity change over the period examined. Finally, model results provide a different perspective on Australian university productivity from past studies, including both the magnitude of productivity change across the sector and relative rankings between institutions. Notable findings in this chapter are also discussed in Moore et al. (2018a) and Moore, Coates, & Croucher (2019).

### **Key Words:**

constructs, construct associations, scope, boundaries, adaptations

## 6.1 Introduction

This chapter presents findings from Stages 3 and 4 of the design research. Findings are generated from assessing the model and its results after being tested on data from nine Asia-Pacific countries and Australia. In the context of the research project, the findings presented in this chapter inform research questions Q1, Q2, and Q3—approaches to understanding higher education productivity, characteristics of an appropriate measurement model, and Australian university productivity trends, respectively. The chapter is organised into four major sections: Asia-Pacific evaluation, Australian data evaluation, model re-design, re-design results. Results from the Asia-Pacific evaluation reveal broad strengths and limitations of the model, as well as diverse examples and evidence for how different international stakeholders understand and conceptualise higher education productivity. Results from the Australian data evaluation supplement international test findings and guide re-design efforts toward increased validity, reliability and relevance for Australian universities.

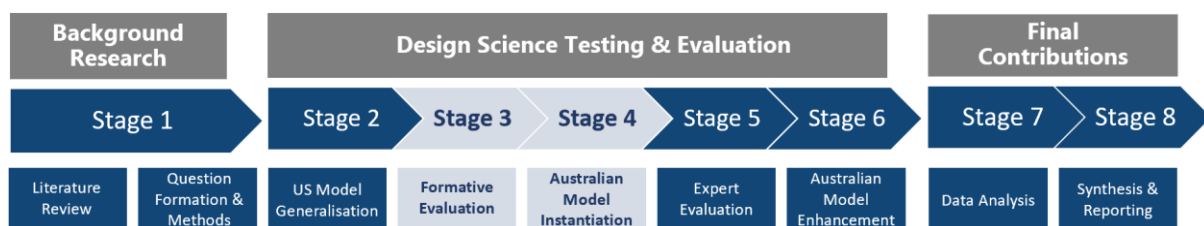


**Figure 13: Research Stages 3 & 4**

The purpose of this chapter is to illustrate and emphasise the level of care that must be taken when adapting productivity measurement tools for higher education contexts. Most measurement techniques were initially constructed to measure productivity phenomena occurring outside the education sector. Sections 6.2 and 6.3 assess all model specifications and their performance in various international higher education contexts. Contextual and practical motivations guide the meticulous vetting process. The objective of the evaluation is not to optimise mathematical protocol or to generate the most robust statistics on a particular set of data—as has been raised as a criticism about the state of productivity scholarship in higher education (De Witte & Sneyers, 2017). Rather, the level of detail in the following two sections is necessary for addressing issues raised in Chapters 2 and 3. Sections 2.6 and 3.5 discuss concerns about how inputs and outputs are chosen for measurement, how they are treated within models, what interests they represent, and how much insight they provide in context. Confronting these scholarly and pragmatic concerns is the value of comprehensively evaluating the model using Williamson and Johanson's (2013) four-point model assessment criteria.

In the interest of framing subsequent discussion, and in helping the reader visualise where the following analysis leads, it is helpful to first list all adaptations made to the measurement

model that derive from the findings of the evaluation described below. These model adaptations are based on (a) empirical observations of the data reported by the participating countries (b) their experiences using the model and (c) how useful the model was for providing insight about individual country contexts. These learnings culminated in six clear options for redesign, and all model adaptations are aimed at improving the capacity of the model to generate targeted insight about the Australian higher education system. The six model adaptations are as follows:

1. Changing the modelling environment from Excel to RStudio
2. Improving estimation of capital service inputs
3. Better representing the time-scale of university production processes
4. Testing ranges of values for non-empirical output value weights
5. Calculating cumulative productivity change across the period
6. Determining thresholds to express ranges of plausible productivity portrayals

## 6.2 Asia-Pacific Model Evaluation

### 6.2.1 Evaluation Methods

The principle methods employed in this chapter reflect those set out by Williamson and Johanson (2013) for design science research evaluation of models. The prime model criteria to be evaluated are (1) model constructs, (2) construct associations, (3) model boundaries, and (4) novelty, revelation and importance. Each criterion is examined by the principle researcher of this dissertation for alignment to the research questions and for suitable model performance in context. The analysis was possible because of direct involvement with the Asia Productivity Organisation (APO) project described in the previous chapter.

The objects of analysis in the Asia-Pacific evaluation are the implementation reports generated by country participants, which compose the Coates (2017a) report on the APO initiative. The current evaluation synthesises findings in the reports with respect to the adapted NRC model used by participants, their results, and the assessment criteria listed above. Reported country results reflect stakeholder experiences after using the adapted NRC model. Study participants were recruited from the APO member states of Cambodia, Fiji, India, Indonesia, Malaysia, Pakistan, the Philippines, Sri Lanka, and Thailand. Participants were selected by the APO because of their influence or access to data in the higher education sector in their respective countries. Table 9 summarises data collection methods and the university datasets analysed by each country.

**Table 9: Data collection by each participating country** (Moore, Coates, et al., 2019).

| Country  | Dataset                                    | Data Collection                                |
|----------|--------------------------------------------|------------------------------------------------|
| Cambodia | 4 public universities                      | Survey completed by participating universities |
| Fiji     | 2 public universities                      | Survey completed by participating universities |
| India    | 82 centrally-funded technical institutions | Government/ministry                            |

|             |                               |                                                           |
|-------------|-------------------------------|-----------------------------------------------------------|
| Indonesia   | 73 public and private HEIs    | Survey completed by participating universities            |
| Malaysia    | 20 public universities        | Government/ministry                                       |
| Pakistan    | 6 public HEIs                 | Public data sources, institution websites, annual reports |
| Philippines | 1,795 public and private HEIs | Government/ministry                                       |
| Sri Lanka   | 14 public universities        | Public data sources, Scopus, Government/ministry          |

The methods and data for each country vary considerably. On one hand, the variety of data renders cross-country comparisons of results nearly impossible. On the other hand—for this evaluation—the ability to assess the model under a variety of conditions affords an ideal opportunity to investigate the limits of the model. Table 9 shows that the sample sizes per country vary considerably. The factors influencing the variability include country size, data availability, and the capacity of the in-country research team. The country datasets also comprised different types of institutions. For example, the four institutions in Cambodia’s sample are full universities. India’s sample includes only technical institutions. Additionally, study participants used different strategies for collecting data. Some participants gathered data via survey instruments that required institutional representatives to provide time-series data on the indicators listed in Appendix A. Other countries with more advanced and integrated higher education databases were able to request data directly from their governments or ministries of education. Other participants gathered data by scraping websites and reviewing public documents, such as annual reports (Coates, 2017a).

## 6.2.2 Model Constructs Assessment

As discussed in Chapter 4, assessing model constructs involves checking for rigorous definitions of constructs and identifying relationships between the constructs and their real-world instances. Country reports provide insight on constructs via their documented choices on model specifications and how those specifications relate to their measurement objectives. Key findings from the output evaluation are listed in Table 10 and explained in more detail below.

**Table 10: Construct assessment findings**

| <b>Finding 1</b>                                                                                                                                                                | <b>Finding 2</b>                                                                                                         | <b>Finding 3</b>                                                                                                                                   |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Overarching model constructs</b> (productivity change, TFP, multiple outputs, joint productivity, and indexing) proved appropriate and adequate to generate desired results. | <b>Input constructs</b> (labour, capital and intermediaries) were logistically successful and generated desired insight. | <b>Output constructs</b> (education and research components) require more rigorous definitions and guidelines for use to generate desired insight. |

The overarching model constructs proved useful for all participating countries. When presented with the opportunity and the support to measure total factor productivity (TFP) change for higher education, representatives of all participating countries of the APO

contributed the necessary time and effort to collect data, analyse results and report findings. Their continued participation for the duration of the initiative bodes well for confidence in the overarching constructs that underpin the productivity measurement model.

Input constructs to the model also proved appropriate across the participating countries. Financial accounting practices across the six participating countries differed, but the prime construct categories of labour, capital and intermediaries were sufficient starting points. Each country was able to disaggregate inputs into these categories. The constructs were successful in representing the value of real inputs that occurred on institutional financial reports. Accounting for multiple input types posed no hurdles logistically or conceptually for the participating countries.

Selecting appropriate data elements for output constructs proved more difficult. Table 11 below summarises the constructs used to represent university outputs across the countries. Both Fiji and Pakistan were able to run the model on comprehensive sets of data, whereas Cambodia Malaysia and The Philippines were more limited in the constructs they portrayed in the model.

**Table 11: Country-by-country output representation decisions**

| Indicator         | Variable               | Cambodia | Fiji | India | Indonesia | Pakistan | Thailand | Sri Lanka | Malaysia | Philippines |
|-------------------|------------------------|----------|------|-------|-----------|----------|----------|-----------|----------|-------------|
| Education outputs | Student Load           | X        | X    | X     | X         | X        |          | X         | X        | X           |
|                   | Coursework completions | X        | X    | X     | X         | X        |          | X         |          | X           |
|                   | Graduate employment    | X        | X    | X     | X         | X        | X        | X         |          |             |
|                   | Learning outcomes      |          | X    |       |           |          |          | X         |          |             |
| Research outputs  | Publications           |          | X    | X     | X         | X        | X        | X         |          |             |
|                   | Citations              |          | X    |       | X         |          |          | X         |          |             |
|                   | Patents                |          | X    | X     | X         |          |          | X         |          |             |
|                   | Research completions   |          | X    |       | X         | X        | X        | X         |          |             |
|                   | Research funds         |          | X    | X     | X         | X        | X        | X         |          |             |

The key finding of the output construct evaluation is that rigorous definitions of model constructs are more important than allowing for flexibility in use of the model. For the well-defined constructs, no problems emerged in reporting and interpreting results. However, the inclusion of ill-defined constructs tended to muddy the waters for interpreting findings and for understanding how well the constructs represented the real-world phenomena of interest. On the education side, individual country definitions of ‘graduate employment’ and ‘learning outcomes’ were least clear. On the research side, ‘patents’ and ‘citations’ were least clear. These constructs showed to be problematic in terms of reporting clarity. Practitioners often provided no specific information about the nature of the data elements. They provided no additional information about how each the data was collected or measured. The consequences are two-fold. First, understanding how the named constructs above contributed to their respective productivity indexes is difficult. Second, making judgements about how these constructs could be improved or included in the next version of the model can now be done on only limited evidence.

Finally, the incorporation of ‘research funds’ or ‘research revenue’ into the model was not straight forward. From an operational standpoint, funding for conducting research is naturally

an input to the research process. Indonesia, for example, chose to portray research funding a key input. Most countries that had data on research income, however, did portray in in their output framework as an outcome indicator of a successful grant application process as a ‘reward’ for the completion of research degrees. Ultimately, some chose to follow the provided guidelines in the input-output framework, and some did not.

### 6.2.3 Construct Associations Assessment

Assessing model construct associations involves checking rigorous definitions of the associations and the nature and appropriateness of the associations. Country reports provide insight on construct associations used in their analyses via their documented choices on how data elements were aggregated and how those calculations relate to their measurement objectives. Key findings from the output evaluation are listed in Table 12 and explained in more detail below.

**Table 12: Construct associations assessment findings**

| <b>Finding 1</b>                                                                                                           | <b>Finding 2</b>                                                                               | <b>Finding 3</b>                                                                  |
|----------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|
| <b>Overarching indexing methods</b> proved effective and sufficient for aggregating data elements for 8 of the 9 countries | The ‘ <b>adjusted credit hours</b> ’ adaptation was not successful for participating countries | The <b>productivity calculation tool</b> itself requires revision and amendments. |

Indexing methods proved useful for eight of the nine participating countries. Seven of the nine countries used Törnqvist indexing with adapted NRC model specifications. Malaysian stakeholders already had expertise using the Laspeyres index—also commonly used for price and cost indexing (Braithwait, 1980)—and chose to implement that technique. Stakeholders in the Philippines were interested in measuring productivity using non-monetary inputs. The Philippines used non-financial proxies for input quality and value, such as institutional totals of accredited programs and non-accredited programs, as well as qualification levels of academic staff. Their interests in measuring productivity diverged from the rest of the group, and they chose to use a regression model instead of index methods.

One consistent deviation from the model across all countries was the calculation of ‘adjusted load’ as a single education output. Different countries track student load and ‘credit hours’ differently. Although ‘credit hours’ from the NRC model was generalised to equivalent full-time student load (EFTSL) in the model, this guideline was still not a feasible for most countries because of the nature of their available data. The calculation of adjusted load would need to be adapted for (a) variations in program length, (b) whether student load information is counted or actual student numbers are counted, and (c) variations in how full-time and part-time students are determined. Countries were thus advised to treat their data proxies for load and completions as separate education constructs and to aggregate them in the same fashion as research constructs. The revised education construct association can be seen below. The following formula takes the variables from the table in Appendix A above and assumes that all are used in the calculation.

$$E_{\Delta t} = \left( \frac{E_{1,t}}{E_{1,t-1}} \right)^{v_{E_1}} * \left( \frac{E_{2,t}}{E_{2,t-1}} \right)^{v_{E_2}} * \left( \frac{E_{3,t}}{E_{3,t-1}} \right)^{v_{E_3}} * \left( \frac{E_{4,t}}{E_{4,t-1}} \right)^{v_{E_4}}, \quad [14]$$

where the baseline assumption is  $v_{E_i} = \frac{1}{k} = \frac{1}{4}$ . And generally,  $\sum_i^k v_{E_i} = 1$ , where  $k$  is the number of education output components used in model.

Construct associations relate also to the procedural, operational aspects of the model. All Törnqvist index calculations were performed using Microsoft Excel. The complexity of the calculation tool provided to APO participants seems to have been prone to error. This is an issue for the utility of the model. Figure 14 demonstrates a likely error in calculation as reported by one country in (Coates, 2017a).

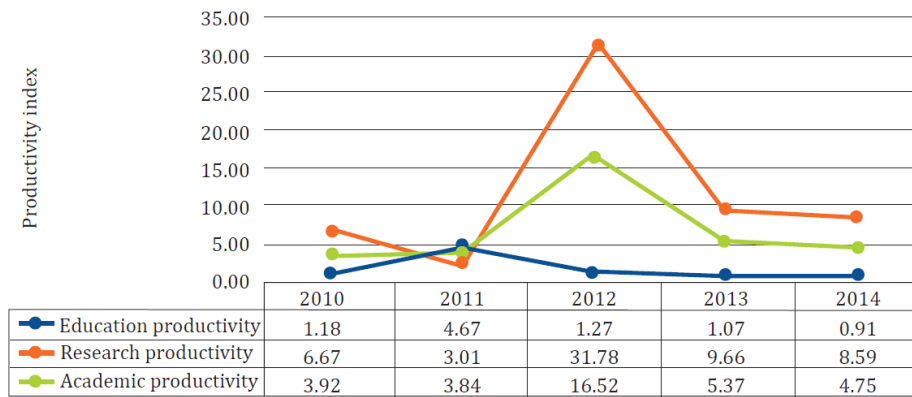


**Figure 14: Reported change figures from Pakistan (Coates, 2017, p. 168)**

Consider the ‘Academic productivity’ data series from Figure 14. If the figures were accurate, they would signify that the institution’s yearly productivity change increased consistently by a factor of five, and in one case by a factor of 15. An examination of the country’s source data in Table 13 shows increases in certain outputs, but the table demonstrates that such extreme trends in  $P_{\Delta t}$  could not be accurate.

**Table 13: Pakistan source data (Coates, 2017, p. 168)**

| Data element         | Unit    | 2010          | 2011          | 2012          | 2013          | 2014          |
|----------------------|---------|---------------|---------------|---------------|---------------|---------------|
| Credit hours         | Hours   | 5,112         | 29,856        | 37,320        | 45,024        | 43,128        |
| Learning outcomes    | Percent |               |               |               |               |               |
| Publications         | Number  | 25            | 12            | 62            | 112           | 181           |
| Citations            | Number  | -             | -             | -             | -             | -             |
| Patents              | Number  | -             | -             | -             | -             | -             |
| Research completions | Number  |               |               | 35            | -             | 57            |
| Research funds       | PKR     |               |               | 410,680,000   | -             | 169,976,294   |
| Labor                | PKR     | 604,196,000   | 693,742,000   | 789,000,000   | 975,000,000   | 1,152,000,000 |
| Capital              | PKR     | 2,000,000,000 | 2,000,000,000 | 2,171,000,000 | 2,611,000,000 | 3,372,000,000 |
| Intermediaries       | PKR     | 846,045,000   | 976,339,000   | 1,018,000,000 | 1,374,000,000 | 1,697,000,000 |
| Total                | PKR     | 3,450,241,000 | 3,670,081,000 | 3,978,000,000 | 4,960,000,000 | 6,221,000,000 |

The error may have been accentuated by the ‘Research completions’ data element fluctuating around zero. The root of the error though was likely unintentional editing or manipulation of the provided Excel calculation tool. The fact that an unintentional edit could have occurred poses a problem for the value of the tool, and a solution is required. The error speaks to a larger problem, however, concerning the interpretation of measurement results. Only in the

event of an error are consistent  $P_{\Delta t}$  figures as high as those in Figure 14 likely to occur. The fact that results were still reported likely indicates that there was difficulty in (a) understanding what the Törnqvist indexes represent, and accordingly in (b) understanding how to extract value from the results. This finding is discussed further in the assessment of novelty, revelation, and importance below.

#### 6.2.4 Model Boundaries Assessment

The assessment of model boundaries explores what phenomena the model represents and what it does not. For the purposes of this study exploring model boundaries involves exploring limitations of the model and its scope for insight. This section focuses on how country reports focused on explicit limitations of the model and where the model was successful in represent productivity in context. Table 14 summarises findings from the evaluation.

**Table 14: Model boundaries assessment findings**

| <b>Finding 1</b>                                                                                                                                        | <b>Finding 2</b>                                                                                        | <b>Finding 3</b>                                                                                                                                                  |
|---------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>The scope for insight of the model</b> is insufficient for addressing certain stakeholder interests in the quality of inputs, processes and outputs. | <b>Current TI specifications</b> may exaggerate or misrepresent true changes in productivity over time. | <b>Approximations of separate research and education productivity</b> suffer from non-empirical judgements about resource allocation to the individual functions. |

The first finding from the boundaries evaluation is the difficulty of the model to capture the quality of higher education services and deliverables. This is not a surprise. Productivity is an inherently mathematical and quantitative concept. However, alternative formulations of a model might be better suited to handling quality indicators. The authors of the Philippines country report, for example, emphasised system-wide needs to focus on quality assurance, enhancement, and quality systems. For this reason, they chose a different set of indicators and a different modelling technique to explore trends in the data.

It is too early to confirm which quantitative techniques may better capture elements of quality. One option used by representatives from The Philippines was to count only instances of outputs deemed to have reached a certain level of quality. Practitioners from the Philippines compared model findings when data was used from all university programs vs. when data was used from only ‘accredited’ university programs. Incorporating quality and value into productivity measures is still under debate. Sullivan et al (2012) recommend that investigations of quality should still be handled in separate, parallel analyses from productivity studies. However, the NRC’s development of the ‘adjusted credit hour’ metric is a direct attempt to incorporate notions of ‘value’ within a university education productivity estimate—rather than accounting for solely quantitative representations of student throughput. Further, the fact Philippines practitioners found utility in accounting for quality in their productivity estimates means that experimenting with quality and value indicators in subsequent versions of the model should be considered. Existing non-empirical inputs to the model may serve as useful tools to address this phenomenon.

The next finding calls into question the efficacy of the current specification of the Törnqvist index to provide reliable estimates of year-on-year university productivity change. Consistent and sometimes drastic spikes in measured productivity of institutions across all participating APO countries highlight limitations in the model to capture true productivity trends. Consider that universities are generally not small, nimble organisations that make drastic changes to their business and service delivery models in short periods of time. Yet, participating APO countries often reported sharp year-on-year turns in productivity change. These sharp turns effectively signify changes to the universities' production functions. The intention of a productivity metric is to reflect the 'technology' that serves to transform inputs into outputs during the time that inputs are being utilised to make the transformation. Consider further that students often spend three years or more in academic programs, and research funding proposals rarely translate directly to research publications within a single calendar year. The time, resources, and energy to produce key education and research outputs are often accumulated, distributed and consumed over more than one year. Hence, when measured university productivity change exhibits unstable patterns from year to year, questions must be asked about how reliable the metric is in representing true productivity.

Investigation into the countries' source data provides some explanation about why spikes in measured productivity change occurred. Annual funding and budgetary changes that are captured by input indexes may serve to exaggerate portrayals of productivity change in ways that do not reflect changes in the 'technology' or key work processes that universities employ to create their outputs. It is true that budget and planning exercises are annual affairs, and they do affect inputs to production, but budgeting and accounting practices are different from exercises in changing or improving productivity. Reports from Cambodia and Pakistan illustrate the point.

Year-on-year fluctuations in government funding and budgetary allocations may significantly affect productivity metrics, but they may not necessarily reflect real changes in institutional productivity. Figure 15 and Figure 16 illustrate how much difference a year can make for the current model's productivity estimates.

Over a five-year period, one Pakistan university received no research funds during three of the five years but received relatively large amounts for two of the five years. Any component indicator that fluctuates around zero in a productivity metric can disproportionately affect final productivity results. Results from Cambodia show figures on capital and intermediate expenditures doubling and then halving in subsequent years. When productivity change metrics isolate changes between only two years, fluctuations in inputs can create large spikes in measured productivity and may represent a significant limitation of the model. For institutions and systems where funding policies and structures change frequently, year-on-year measures may not give an accurate or reliable portrayal of productivity change.

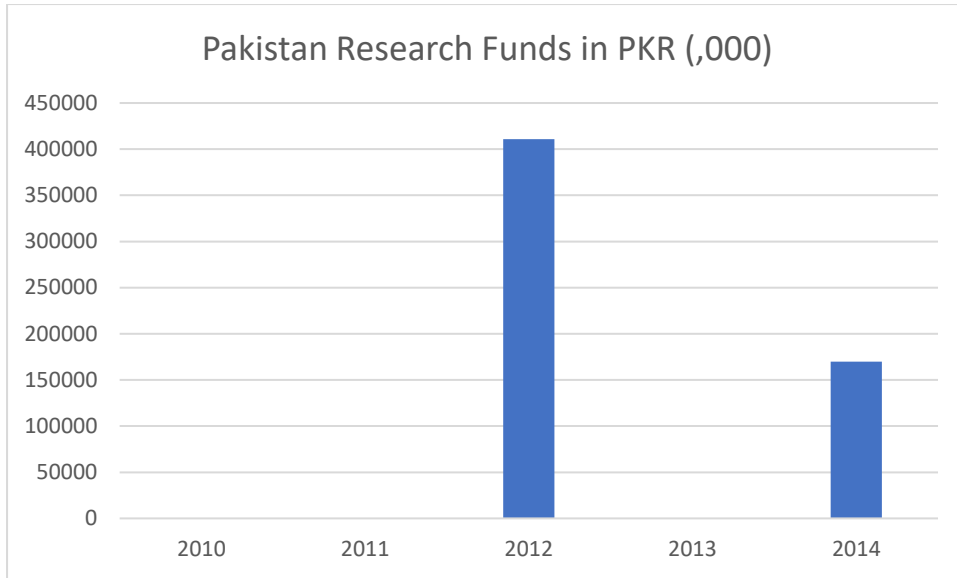


**Figure 15: Fluctuations in Pakistan's source data**

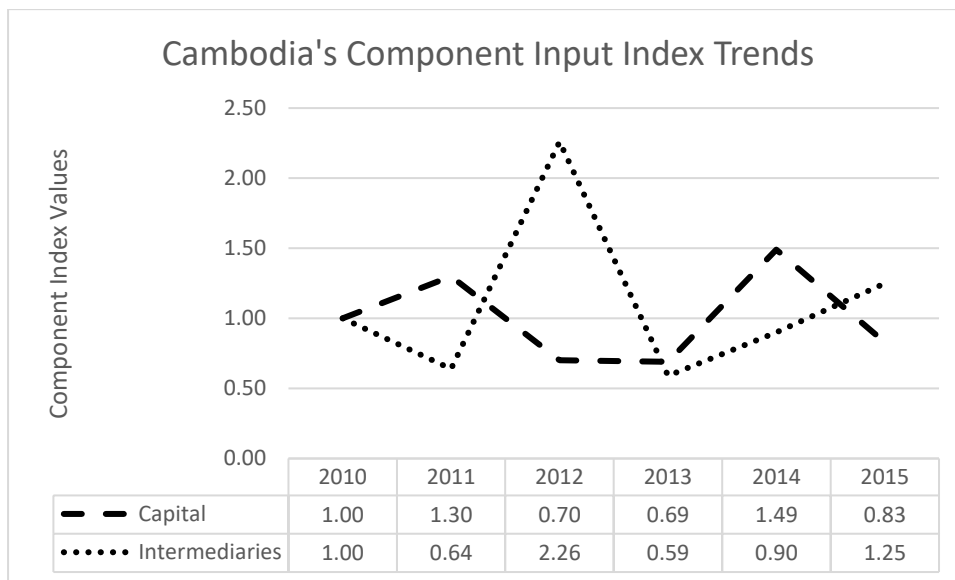


**Figure 16: Fluctuations in Cambodia's component input indexes**

The last finding from the evaluation of model boundaries regards limitations of the model to provide separate estimates of education and research productivity. A baseline assumption of the model is that inputs are split evenly between education and research functions. As discussed in Chapter 3, this is a common assumption for production functions that measure joint productivity of outputs. A derivative calculation to estimate separate productivity of academic functions can thus be made with an even split of inputs to each function. The calculation can be performed by taking  $\frac{E_{\Delta t}}{(\frac{1}{2}X)_{\Delta t}}$  and  $\frac{R_{\Delta t}}{(\frac{1}{2}X)_{\Delta t}}$ .

A more valid estimate of productivity for the separate academic functions would require institutional financial reports that account for resource allocations to the academic functions. In effect, the practitioners would need to know input estimates that satisfy the equation  $X_A =$

$X_E + X_R$  on a yearly basis, where  $X_A$  is total inputs to both academic functions,  $X_E$  is education inputs, and  $X_R$  is research inputs. Indonesian country participants were able to make crude estimates of  $X_E$  and  $X_R$  because of their unique accounting system. Other country participants, however, still found utility in checking estimates of productivity change for each academic function using the method described in the previous paragraph. Figure 17 provides an illustration of how the model allows for productivity index comparisons between education, research, and composite academic productivity. The accuracy of the separate education and research productivity indexes without accurate input apportionment to each function, however, is questionable.

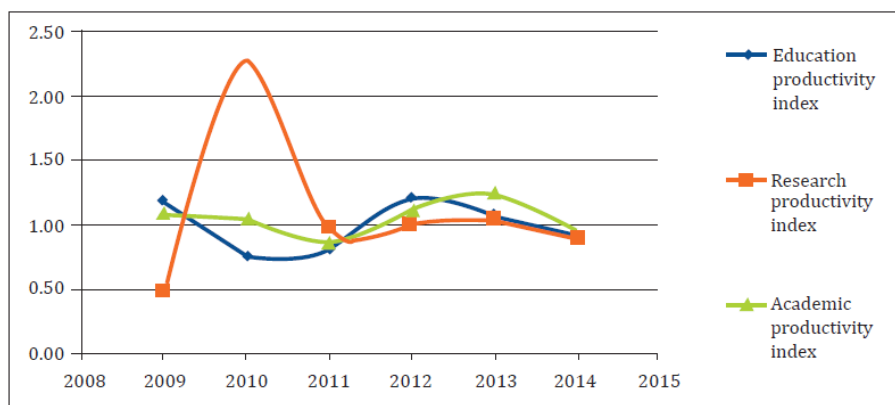


Figure 17: Indonesia education, research and composite indexes (Coates, 2017a, p. 106)

### 6.2.5 Revelation, Novelty, Importance Assessment

The assessment of revelation, novelty and importance focuses on phenomena revealed by the model or uses of the model that were not previously known. If phenomena were already known, it focuses on where the model provides a deeper understanding. Evidence for this evaluation derived largely from conclusion and discussion sections of APO country reports. Key findings from the novelty, revelation and importance evaluation are summarised in Table 15.

Table 15: Revelation, novelty and importance assessment findings

| Finding 1                                                                                   | Finding 2                                                                                    | Finding 3                                                                                                                                             |
|---------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>The model revealed</b> useful information for both national policy and funding contexts. | <b>A novelty</b> of the indexing method is its accessibility and readiness for decomposition | <b>Regarding importance,</b> measurement results instilled a sense of urgency among participants about the value of improving measurement techniques. |

A key revelation from testing the measurement model is that it has shown to be informative under a variety of national contexts and useful under different circumstances of data availability. The country report from India gave results for single institutions. The Thailand report used a larger dataset and aggregated universities into groups based on institutional size.

The Cambodian report used the model to provide system-level productivity indicators by aggregating data from all institutions. Finally, the Fiji report discussed how the model could also be used to examine departments within a university. The formative evaluation of the model instils confidence that further development for an Australian instantiation of the model could provide new insight about the Australian system at multiple levels of operation and provision.

Country reports emphasised the potential of productivity change analysis for generating key information about system and institutional functioning and provision. In the context sections and concluding remarks of the country chapters, a prime impetus often cited for interesting measuring productivity change was the growth of higher education in the participating countries. Participating countries indicated that they took part in the APO initiative because they were interested in learning how best to provide affordable higher education to fast growing markets. Understanding productivity helps practitioners and stakeholders understand the implications of changes in system-level governance, infrastructure development, regulation, and changes in key work processes. Examples include establishing new ministries, implementing new legislation, appropriately resourcing institutions, accrediting institutions, and learning about the operations of public vs. private universities. Rapid and wide-ranging changes require comprehensive and dynamic indicators to gauge performance trends.

A Novelty of the TI aggregation technique is its provision for decomposition of model components. It allowed for analysis of separate trends in research and education productivity. The inclusive nature of the model allows for results that can provide an important baseline for conversation about key performance trends. For example, results from some countries showed concurrent productivity gains in both research and education. Results from other countries showed divergent trends, where productivity gains for one function correlated with productivity losses for the other function. Such information can provide baseline information for discussions about the teaching-research nexus of an institution or system. Further information and deeper analysis would be required to make any conclusions about a thriving teaching-research nexus, or alternatively, a zero-sum game between functions competing for resources. However, different institutions and systems have different policies, regulations, and key work processes, and analyses such as this can produce evidence for how those policies, regulations, and processes translate into results for the prime functions that universities perform.

A final revelation concerning the importance of findings is a clear demonstration of how urgent it is to develop and improve productivity indicators for the higher education context. Common theme raised in the country reports was that of Institutional management and autonomy. Effective management of institutional resources and operations was cited as a relentless challenge by authors of the reports. Performance indicators can act as both lead and lag indicators for performance, as well as both carrots and sticks in the process of improving transparency and accountability. Only accurate and reliable metrics that us appropriate observations and evidence can inform quality management and decision-making. Institutional autonomy to manage and direct resources and key work processes varies from country to country. Leveraging and optimising autonomy in the domains where it is afforded means extracting as much useful information as possible from available data and understanding where further data collection is most important.

Higher education productivity indicators differ from solely financial indicators, and they need further development. Higher education productivity often incorporates financial information, but the purpose of productivity analysis is to understand the linkages between budgets, resource management strategies, key deliverables, and strategic outcomes. Productivity metrics do not paint a complete picture of higher education institutional functions and operations, but they reveal narratives about performance that singular indicators cannot provide in isolation. APO country reports stress that refining the practice of productivity measurement can help provide a new and important type of information for improving the provision of higher education.

### **6.2.6 Asia Pacific Evaluation Conclusions**

A key purpose of the evaluation is to build a foundation of evidence for identifying areas for actionable revision of the model. Before discussing technical implications, however, the revelation, novelty and importance assessment revealed additional big-picture insights about productivity measurement in context. Participants reported that their productivity change estimates helped to frame issues such as industry engagement practices, adoption of new technologies, effects of upstream secondary education systems, and differences between public and private institutions. Country stakeholders justified their participation in the APO initiative because of a desire to provide more affordable higher education to fast growing markets. They showed concern for how large-scale changes occurring within systems affect institutional results. Country reports cited disruptive system-level activities, such as establishing new ministries, implementing new legislation and regulations, providing ministries with extra human and financial resources, and establishing funding and accreditation bodies. These contextual issues illustrate how prior understandings of performance and productivity can both shape analyses and provide key information about contemporary issues. The accessibility of data in participants' respective countries limited what they could analyse. Specific performance improvement interests within their systems, however, drove data selection and treatment priorities with respect to what data could be accessed.

Use of the model by country practitioners revealed that productivity change estimates are not always easy to interpret. The  $P_{\Delta t}$  Törnqvist indexes themselves represent trends in productivity. Thus, when  $P_{\Delta t}$  figures are presented and visualised in standard line charts, their meaning could be skewed. Trends in line charts are often understood to be represented by the slopes of the lines in the illustration, not by the individual data points. This complexity likely contributed to at least one participant country reporting confirmed bogus results. Subsequent re-designs of the model should focus on improving the productivity change analysis to make results more user friendly and more human interpretable.

With respect to model specifications, the evaluation revealed several priority areas for model improvement in subsequent redesign. For model input constructs, 'graduate employment' and 'learning outcomes' must be addressed. Variation in data availability, differences in measurement, and lack of clarity in reported methods, proved to confound results rather than render the model more flexible. The same issue exists for the output constructs of 'citations' and 'patents'. The Australian evaluation below provides further insight into how addressing model constructs can be accomplished. Details are provided in section 5.4.

Model construct associations and boundaries must also be addressed in subsequent redesign. Both the concept and calculation of ‘adjusted load’ is not as straight forward as anticipated, and must be handled with care in the Australian instantiation of the model. How the model handles education and research outputs separately to provide alternative perspectives on performance needs to be addressed. The availability of appropriate empirical data to separately examine academic functions poses problems for the validity of results.

## **6.3 Australian Data Evaluation**

### **6.3.1 Evaluation Methods and Measurement Results**

The evaluation of the productivity measurement model on Australian data supplements the findings of the international assessment. The Asia-Pacific evaluation provided general information about how the model could be improved, and the Australian data evaluation provides additional information to help direct the model redesign processes for Australian specific contexts. After being tested on Australian data, the model and its results are evaluated via the same four model assessment criteria as above.

The deployment of the adapted NRC model on Australian data includes indicators from 38 Australian universities. The analysis examines the period of 2007-2013. The period was chosen for comparative value to prior productivity change analyses performed on Australian data. The ability to examine model results and compare them directly to prior studies allows for additional potential insight on model performance. Specifications for the model used on Australian data are the same as those explicated in sections 5.3 and 5.4 of the previous chapter. All data was collected from public sources through the Australian Department of Education and Training, as indicated in the previous chapter. Each data element was collected in a consistent manner across all institutions. For easy reference, Table 16 restates the list of data elements used in the model.

Measurement results from the model for all Australian public universities with complete data across the period are shown in Table 17. The Törnqvist indexes  $P_{\Delta t}$  for each university over the period provide the foundation for model evaluation. A summary of measurement results indicates that the median rate of productivity change for each year sat in the range from  $P_{\Delta t} = 0.95$  to  $P_{\Delta t} = 1.01$ . The year 2013 appears to have been the best year across the system for productivity change. No single institution consistently achieved the highest  $P_{\Delta t}$  values or the lowest  $P_{\Delta t}$  values across the period. However, universities that exhibited some of the highest scores in certain years during the period also tended to exhibit some of the lowest scores in other years during the period. The behaviour signifies tendencies for some institutions to be prone to large spikes and rebounds in their productivity change estimates. Further it was rare for any institution not to experience a single year of productivity growth. Most institutions experienced at least two years of positive productivity change across the period.

**Table 16: Adapted NRC model input-output framework for Australian data**

| Input Type     | Construct              | Variable       | Data element                            |
|----------------|------------------------|----------------|-----------------------------------------|
| Labour         | Academic staff         | L <sub>1</sub> | Academic staff salary and benefits      |
|                | Non-academic staff     | L <sub>2</sub> | Non-academic staff salary and benefits  |
| Capital        | Land                   | K <sub>1</sub> | Land capital services                   |
|                | Buildings              | K <sub>2</sub> | Buildings capital services              |
|                | Equipment & other      | K <sub>3</sub> | Equipment and other capital services    |
|                | Repairs & maintenance  | K <sub>4</sub> | Repairs and maintenance                 |
| Intermediaries | Administration & other | I <sub>1</sub> | Administration and other expenses       |
|                | Grants & scholarships  | I <sub>2</sub> | Grant and scholarship expenses          |
| Output Type    | Construct              | Variable       | Data element                            |
| Education      | Student Load           | E <sub>1</sub> | Number of full-time coursework students |
|                | Coursework completions | E <sub>2</sub> | Number of coursework graduates          |
| Research       | Publications           | R <sub>1</sub> | Number of weighted publications         |
|                | Research completions   | R <sub>2</sub> | Number of research graduates            |
|                | Research income        | R <sub>3</sub> | Amount of research income               |

**Table 17: Adapted NRC model results on Australian data**

| Institution                            | 2007 | 2008  | 2009  | 2010  | 2011  | 2012  | 2013  |
|----------------------------------------|------|-------|-------|-------|-------|-------|-------|
| Australian Catholic University         | 1    | 0.944 | 1.032 | 1.097 | 0.948 | 0.794 | 1.067 |
| Australian National University         | 1    | 1.022 | 1.003 | 1.029 | 0.956 | 0.983 | 1.020 |
| Central Queensland University          | 1    | 0.997 | 0.925 | 0.993 | 1.024 | 0.920 | 1.205 |
| Charles Darwin University              | 1    | 0.897 | 0.918 | 0.979 | 1.119 | 0.947 | 0.877 |
| Charles Sturt University               | 1    | 0.918 | 0.829 | 1.122 | 0.907 | 0.895 | 1.049 |
| Curtin University of Technology        | 1    | 0.878 | 0.973 | 0.996 | 0.964 | 0.946 | 0.951 |
| Deakin University                      | 1    | 0.996 | 1.010 | 0.910 | 0.922 | 1.052 | 0.946 |
| Edith Cowan University                 | 1    | 1.031 | 0.923 | 1.003 | 0.988 | 0.923 | 1.033 |
| Federation University Ballarat         | 1    | 0.905 | 1.014 | 0.941 | 0.965 | 0.987 | 0.878 |
| Flinders University                    | 1    | 0.984 | 0.886 | 0.982 | 1.005 | 0.974 | 0.950 |
| Griffith University                    | 1    | 0.937 | 0.980 | 1.012 | 0.986 | 0.914 | 0.982 |
| James Cook University                  | 1    | 0.883 | 1.068 | 0.947 | 1.021 | 0.939 | 0.999 |
| La Trobe University                    | 1    | 0.887 | 1.080 | 0.996 | 1.016 | 0.946 | 1.029 |
| Macquarie University                   | 1    | 0.940 | 0.953 | 0.934 | 1.032 | 0.979 | 0.959 |
| Monash University                      | 1    | 0.963 | 0.968 | 0.931 | 1.045 | 0.948 | 1.052 |
| Murdoch University                     | 1    | 0.869 | 1.177 | 0.981 | 1.021 | 0.976 | 0.960 |
| Queensland University of Technology    | 1    | 0.959 | 0.934 | 0.968 | 0.982 | 1.004 | 0.975 |
| RMIT University                        | 1    | 1.004 | 1.003 | 0.932 | 1.048 | 0.959 | 1.047 |
| Southern Cross University              | 1    | 0.940 | 0.928 | 0.915 | 0.963 | 0.982 | 0.979 |
| Swinburne University of Technology     | 1    | 1.073 | 0.935 | 1.101 | 0.943 | 0.964 | 1.150 |
| The University of Melbourne            | 1    | 0.969 | 0.989 | 0.926 | 1.007 | 0.954 | 0.988 |
| The University of New England          | 1    | 0.922 | 0.976 | 0.992 | 0.929 | 0.975 | 0.975 |
| The University of New South Wales      | 1    | 0.919 | 1.043 | 1.006 | 0.928 | 0.974 | 1.009 |
| The University of Newcastle            | 1    | 0.945 | 0.977 | 0.973 | 0.900 | 1.019 | 1.028 |
| The University of Notre Dame Australia | 1    | 1.070 | 0.790 | 1.199 | 0.710 | 0.935 | 1.253 |
| The University of Queensland           | 1    | 0.925 | 1.013 | 0.927 | 0.975 | 1.006 | 1.044 |
| The University of Sydney               | 1    | 0.953 | 0.930 | 1.013 | 0.973 | 0.967 | 1.020 |

|                                     |   |       |       |       |       |       |       |
|-------------------------------------|---|-------|-------|-------|-------|-------|-------|
| The University of Western Australia | 1 | 0.997 | 0.971 | 1.030 | 1.007 | 0.982 | 0.962 |
| University of Adelaide              | 1 | 0.959 | 0.963 | 0.947 | 0.923 | 1.008 | 0.987 |
| University of Canberra              | 1 | 1.027 | 0.856 | 1.136 | 0.995 | 0.943 | 0.958 |
| University of South Australia       | 1 | 0.963 | 1.063 | 1.010 | 0.902 | 0.926 | 1.025 |
| University of Southern Queensland   | 1 | 0.895 | 1.094 | 0.918 | 0.988 | 1.077 | 0.902 |
| University of Tasmania              | 1 | 0.963 | 1.002 | 0.985 | 0.960 | 0.995 | 1.020 |
| University of Technology, Sydney    | 1 | 0.935 | 0.975 | 0.953 | 0.928 | 0.986 | 1.049 |
| University of the Sunshine Coast    | 1 | 0.985 | 1.127 | 0.907 | 0.909 | 1.091 | 0.975 |
| University of Western Sydney        | 1 | 0.909 | 0.948 | 0.963 | 0.961 | 0.943 | 1.011 |
| University of Wollongong            | 1 | 0.923 | 1.027 | 0.930 | 0.953 | 0.979 | 1.034 |
| Victoria University                 | 1 | 1.054 | 0.997 | 0.983 | 0.919 | 0.987 | 1.075 |
| <i>Min</i>                          |   | 0.869 | 0.790 | 0.907 | 0.710 | 0.794 | 0.877 |
| <i>Median</i>                       |   | 0.949 | 0.976 | 0.981 | 0.965 | 0.974 | 1.010 |
| <i>Max</i>                          |   | 1.073 | 1.177 | 1.199 | 1.119 | 1.091 | 1.253 |

The intention of the evaluation is not to provide deep interpretive analysis of results. Results will change as the model is re-designed. The intention of the results summary is to produce sufficient knowledge about how the model performed with the data, so that it can be evaluated using the four model assessment criteria.

### 6.3.2 Model Constructs Assessment

The construct assessment verifies rigorous definitions of constructs and identifies relationships between the constructs and their real-world instances. Publicly available information on all data elements used in the model serves as evidence in this assessment. Key findings from the output evaluation are listed in Table 18 and explained in more detail below.

**Table 18: Construct evaluation findings**

| <b>Finding 1</b>                                                                                                                                                                                       | <b>Finding 2</b>                                                                                                                                                                                                          |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>The reliability of source data</b> allows for more confidence in measured results, but the ‘weighted publications’ indicator lacks a precise empirical rationale for representing scholarly output. | <b>A reduced number of input and output constructs</b> may limit the scope of university performance portrayals, but instead allows for more targeted and endogenous productivity portrayals for individual institutions. |

All input and output constructs included in the model for Australian data are backed by data elements that have been rigorously defined. Each data element represents real world phenomena. The Australian department of education and training (DET) collected and reported all data in a rigorous manner over the period examined. The education data elements come from the Australian uCube database. The research data elements come from the Australian Higher Education Research Data Collection (HERDC) repository. The figures from uCube and the HERDC repository are used within the system for institutional decision-

making and funding. They were chosen because they have been *de facto* important and influential figures for all institutions in the system.

The HERDC figure, ‘weighted publications’, was computed by the DET over the period examined with different value weights assigned to different types of publications. The weights signify judgements about the relative impacts of different publication types, such as scholarly articles, book chapters, and presentations. ‘Weighted publications’ figures provide a research analogue to the ‘adjusted load’ calculation within the education construct. Just as ‘adjusted load’ seeks to account for the extra value add of degree completion, ‘weighted publications’ seeks to account for the extra value add of different research publication types. The metric is not perfect, but it was standard to the system during the period examined, and it was used in funding formulas for research grants. It has therefore been a key research output for Australian public universities.

Data elements omitted from the model for Australian data include ‘graduate employment’, ‘learning outcomes’, ‘citations’, and ‘patents’. These constructs are not backed by robust data in the system that have been defined and collected by public authorities as rigorously as has been done for the data elements discussed above. The inclusion of additional indicators for the international assessment of the model provided an opportunity to consider whether additional data collection from non-Australian government sources might be worth the effort in the Australian case. This remains an open possibility for the model. The measurement model needs further revision first, however, before including potentially less robust constructs that may confound results in unexpected ways. Limiting the number of output constructs to the model limits the scope for what the model can represent. This is not necessarily undesirable, though, and the issue is discussed further in the assessment of model boundaries.

### 6.3.3 Construct Associations Assessment

The construct associations assessment involves checking for rigorous definitions of the associations and the nature and appropriateness of the associations. Institution by institution measurement results provide insight on construct associations and how they represented input-output dynamics of Australia’s universities. Key findings are provided in Table 19 below.

**Table 19: Construct associations assessment findings**

| <b>Finding 1</b>                                                                                  | <b>Finding 2</b>                                                  | <b>Finding 3</b>                                                                                                                              |
|---------------------------------------------------------------------------------------------------|-------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|
| <b>The ‘adjusted load’ construct association</b> performed reliably among Australian institutions | <b>Capital services estimation</b> needs more empirical grounding | <b>Research and education output value weights</b> need more suitable treatment and rationale, given the diversity of Australian institutions |

The ‘adjusted load’ adaptation for the model from the NRC model’s ‘adjusted credit hours’ worked well for Australian data. Although deep analysis of measurement results is reserved for the final version of the model, the ‘adjusted load’ association between EFTSL and coursework completions holds the potential to provide insight about how universities are

prioritising productivity improvements: through increasing completions rates or through increasing enrolment rates, or both.

The non-empirically based capital service factor for estimating capital services is not appropriate for the Australian context. While it allowed for initial testing of the of the model across multiple international contexts, a fixed capital service factor of 0.05 in the Australian context remains an unreliable figure among the more rigorously defined constructs and construct associations within the model. The function of the capital service factor is to ensure that the effects of changes in capital for the composite input index  $X_{\Delta t}$  are not exaggerated. Total book value of non-current capital does not represent a yearly expense or input. As described in the previous chapter, input components in a Törnqvist index are weighted based on their respective value shares from total inputs. Although using the capital service factor does not make a difference in the measured yearly change in the capital component, it does make a difference for the magnitude of its value weight in the calculation. A more accurate and reliable calculation would be empirically based and be allowed to vary according to yearly dynamics. Further revisions of the model need to improve upon this crude estimate of capital services.

The  $v_R = 0.5$  and  $v_E = 0.5$  output value weights used to aggregate education and research outputs into a single output change index in the model performed well on Australian data. However, while the equitably assigned value weights for the respective academic functions are aligned with historical academic values and ideals within the system (Dekeyser, Watson, & Baré, 2016), exploring how to more appropriately assign the value weights based on individual institutional contexts remains an important feature of the model to improve upon for increased validity and utility.

#### 6.3.4 Model Boundaries Assessment

The assessment of model boundaries explores limitations of the model and its scope for insight. This section focuses on institutional results generated by the model and how it could better capture input-output dynamics within the Australian system. Table 20 summarises findings from the evaluation.

**Table 20: Model boundaries assessment findings**

| Finding 1                                                                                      | Finding 2                                                                                                             | Finding 3                                                                                         |
|------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------|
| <b>Year-on-year change measurements</b> may not reflect true changes in production technology. | <b>Better accounting for the value of inputs and outputs</b> could help with the validity and reliability of results. | <b>The lack of absolute productivity estimates</b> leaves open room for interpretation of results |

The model boundaries evaluation focuses on the scope of the model and its limitations. Similar to what was found during the international assessment, the current specifications of the Törnqvist index may present limitations in portraying the dynamics of an institution's true production technology. Results from the model summarised above indicate large spikes and rebounds in measured productivity change for many of the institutions in the sample. Figure 18 illustrates the situation for three universities. Noisy results such as these have little inherent value or utility. Not all universities demonstrate such dramatic spikes over the

period, so some comparative value can be derived from noisy results. Dramatic changes in measured productivity likely point to different institutional circumstances than those of institutions showing more linear productivity change trends. Interpreting the results from Figure 18 by themselves, however, is difficult. It is unlikely that the institutions' true key work processes and production technologies underwent yearly changes as dramatic as those portrayed in the chart. Exploring how to better represent year-on-year productivity change in the model and developing tools to understand the drivers of such behaviour in results will be important for further redesign.

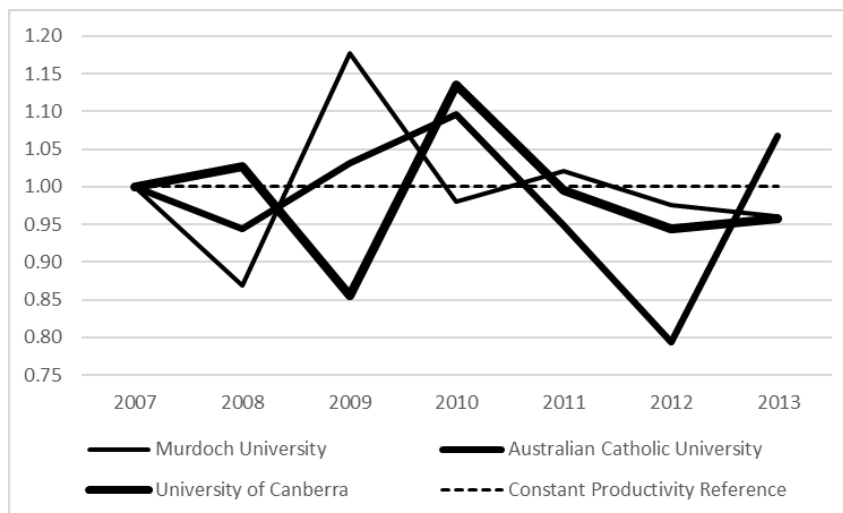


**Figure 18: Spikes in measured yearly productivity change**

Eliminating more loosely defined outcome indicators, such as ‘graduate employment’ allows for development of a more robust model, fit for purpose it its intended representations and portrayals of productivity change. Accounting for the value of outputs, however, remains an important challenge and feature of the model. Incorporating indicators of value helps to distinguish between (a) purely quantitative, context-free productivity portrayals and (b) productivity portrayals that better serve as lead indicators for institutional achievement of desired outcomes in context. As discussed above, the current model includes indicators of value on both the education and research sides of the equation. Exploring options for improvement, however, remains a priority for model redesign.

Finally, measuring only productivity change represents a departure in scope when compared to prior productivity analyses on Australian data. As discussed in Chapter 3, Productivity change findings from prior studies on the Australian system use methods that also estimate frontier efficiencies and proxies for institutions’ absolute productivities. Determining the implications, costs, and benefits of a pure productivity change focus remains critical for subsequent interpretations of results and model redesign. Some initial insights are presented in the assessment that follows.

### 6.3.5 Revelation, Novelty, Importance Assessment

The assessment of revelation, novelty and importance focuses on phenomena revealed by the model or uses of the model that were not previously known. If phenomena were already known, it focuses on where the model provides a deeper understanding. Key findings from

the novelty, revelation and importance evaluation are summarised in Table 21: Revelation, novelty and importance assessment findings.

**Table 21: Revelation, novelty and importance assessment findings**

| <b>Finding 1</b>                                                                                                                                             | <b>Finding 2</b>                                                                                                                                                                                  |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Measurement results revealed a different perspective</b> on higher education productivity in Australian from those of past studies on the same phenomena. | <b>Regarding importance, the variation in measured results</b> per institution likely indicates true contextual differences in institutions, which merit closer examination in subsequent models. |

Running the model on Australian data revealed a new perspective on productivity change for the Australian higher education system. Results appear to stand in contrast to a prior study of institutional productivity change conducted on data from the same period. Moradi-Motlagh et al. (2016) reported an average sector-wide productivity growth of 15.7% over the period from 2007 to 2013 using a DEA calculation model. The results from the model above however, indicated median yearly productivity change rates below a value of 1.0 for five out of the six yearly productivity change intervals during the period. Although the current model does not show a total change from 2007 to 2013, it can be inferred that the average productivity change for the sector over the period is likely not above 1.0, signally negative productivity change. Deeper analysis is required to confirm this result—and to make a value judgment about whether a measured positive or negative change on this data is a good or bad result. The importance of the revelation is that different techniques for calculating university productivity change may render inconsistent results. Subsequent analysis and redesign of the model must address this issue.

Measurement results also reveal the diversity of Australian universities in terms of their productivity processes and trends. Results likely indicate diverse institutional contexts under which the universities operate. Such institutional diversity demands techniques that can unpack the data being used to compute the metrics. The model and its results will only reach their potential for value and utility if they can and speak to the multiple interests and contexts intertwined with the procedural and operational aspects of productivity.

### **6.3.6 Australian Data Evaluation Conclusions**

The evaluation of the model on Australian data contributed key insights for model redesign. For inputs, the estimation of capital services in the Australian context needs to be improved and backed by more empirical techniques. This will affect both constructs and construct associations for the capital input variables. How the model treats inputs in general needs to be improved as well. As revealed in both Step 1 and Step 2 of the evaluation, fluctuations in funding and resource allocation can results in spikes measured productivity that may not necessarily represent true shifts in production technology and key work processes of the institution.

From a contextual point of view, in the model boundaries evaluation, the issue was raised that the model needs improved techniques to account for output value. Likewise, from a technical point of view, the construct association evaluation revealed that the model needs improved

techniques for assigning  $v_R$  and  $v_E$  value weights. These findings are inextricably linked. Combined with the joint finding from both Step 1 and Step 2 that the derivative calculation  $\Delta P_{\Delta t}$  does not provide any additional insight, there is an opportunity to redesign this derivative calculation to be used in conjunction with  $v_R$  and  $v_E$  figures. A better designed supplementary calculation could increase interpretive value of results address many of the evaluation findings from both steps. Details are provided in the following chapter.

Figure 19 highlights in orange the elements of the model to be revised based on findings from the evaluation. First, how the model treats all inputs needs to be addressed. Second, the variables for capital and capital service calculation need revision. Third, an improved method for determining and assigning value weights to academic outputs needs to be developed. Fourth, a characterisation of the data that helps with interpretation of results needs to be developed.

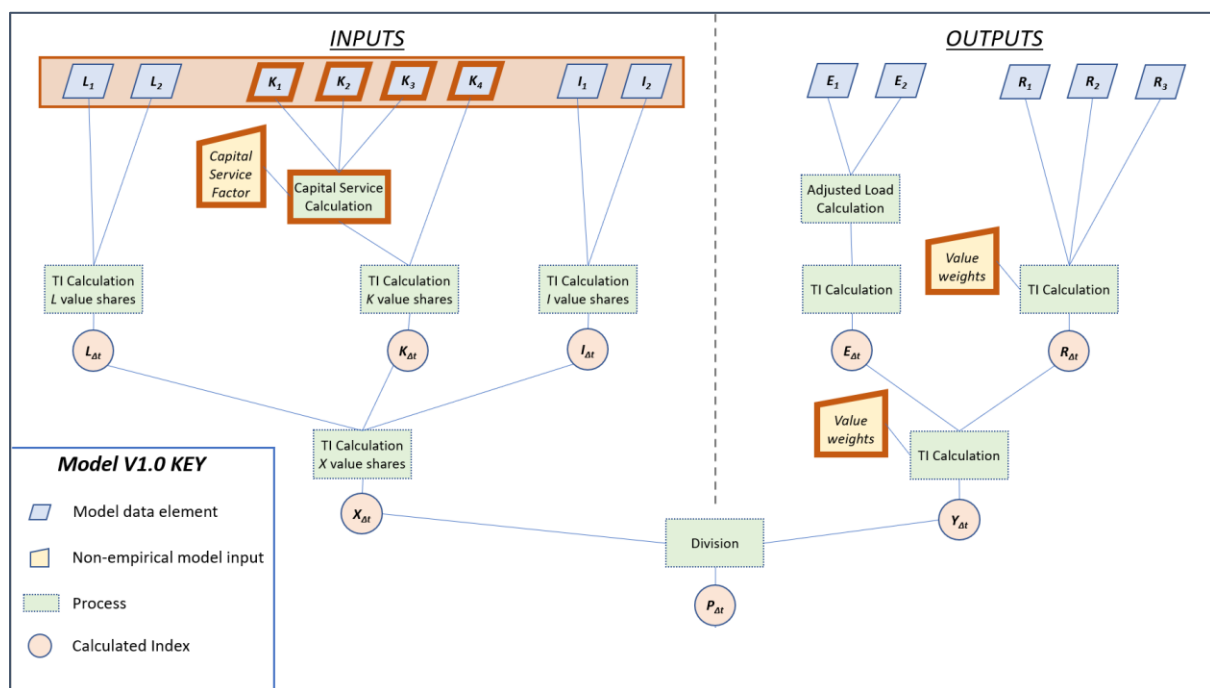


Figure 19: Measurement model schematic with highlighted elements for revision

## 6.4 Australian Model Instantiation

### 6.4.1 Model Redesign

This section synthesises all findings from the model evaluation discussed above. It directs insight gained from model assessment toward model improvement and redesign. The section defines and justifies all construct and construct association redesign specifications for the Australian model instantiation. The purpose of the Australian instantiation is to incorporate broad findings from the international generalisation of the NRC model to produce improved productivity portrayals in the Australian context. The motivation is for the redesigned model

and its results to be evaluated by key stakeholders and experts working within the Australian system. The stakeholder and expert evaluation is described in the next chapter.

Restating details from the introductory section of this chapter, six adaptations to model constructs and construct associations are identified to redesign the second version of the model. The adaptations include:

1. Modelling environment change
2. Direct capital service estimation
3. Input ‘smoothing’ for better production time representation
4. Output value weight range accommodation
5. Cumulative productivity change calculation
6. Productivity threshold estimation

The adaptations are discussed in sequence in the sections that follow. Each adaptation is presented in terms of how it addresses the findings from the evaluation above. Table 22 gives the updated input-output framework for the redesigned model. Figure 20 provides a visualisation and schematic of the model with redesigned elements highlighted in red.

**Table 22: Redesigned model input-output framework**

| Construct Type | Construct                  | Variable | Data elements                           |
|----------------|----------------------------|----------|-----------------------------------------|
| Input          | Labour                     | $L_1$    | Academic staff salary and benefits      |
|                |                            | $L_2$    | Non-academic staff salary and benefits  |
|                | Capital Services           | $K_1$    | Depreciation and amortisation           |
|                |                            | $K_2$    | Repairs and maintenance                 |
|                |                            | $K_3$    | Interest payable                        |
|                | Intermediaries             | $I_1$    | Grants and scholarships                 |
|                |                            | $I_2$    | Administration and other expenses       |
| Output         | Equivalent FT Student load | $E_1$    | Number of full-time coursework students |
|                | Coursework completions     | $E_2$    | Number of coursework graduates          |
|                | Publications               | $R_1$    | Number of weighted publications         |
|                | Research completions       | $R_2$    | Number of research graduates            |
|                | Research income            | $R_3$    | Amount of income for research           |

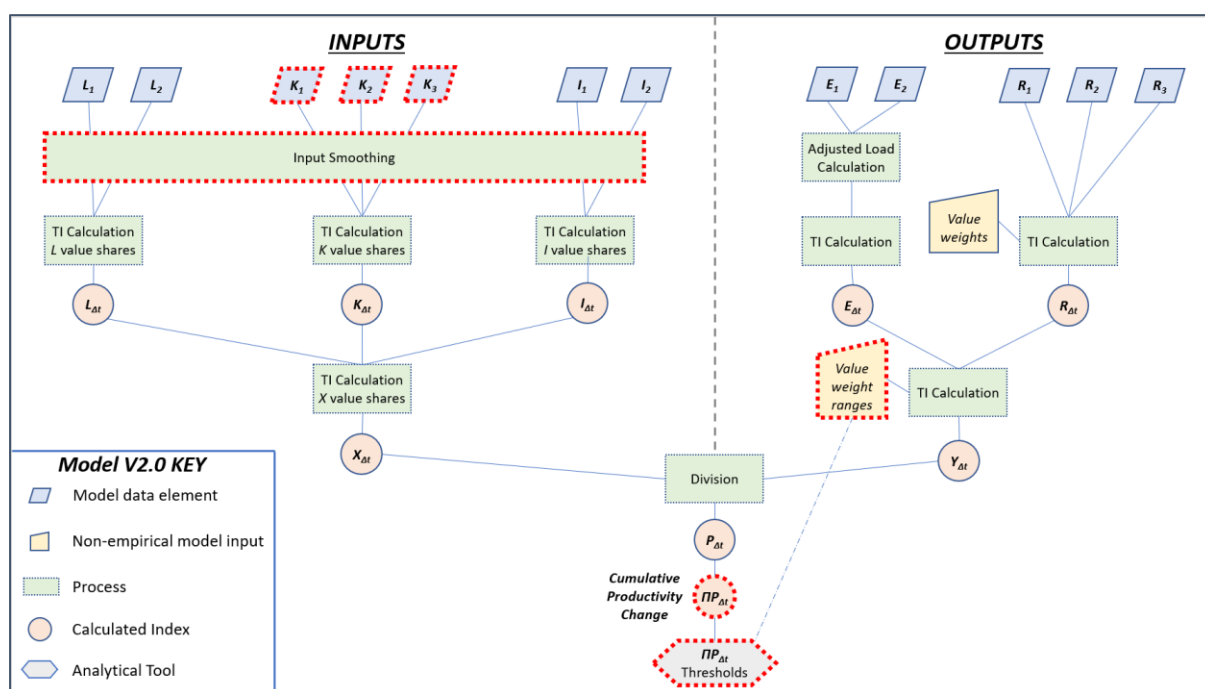


Figure 20: Model redesign schematic

## 6.4.2 Modelling Environment Change

The first iteration of the model used Microsoft Excel to perform all calculations with institutional data. The re-design uses R and R Studio for all calculations and data modelling. The reasons for the modelling environment change are three-fold. (1) As shown in Asia-Pacific evaluation above, the initial Excel model showed that it may be prone to error. Inadvertent editing of the excel file by one of the APO reporting countries likely led to erroneous measurement results for that country. (2) As the model undergoes further redesign, the R environment places the mathematical model as the central item for revision and analysis. In Excel, the central items for analysis are individual cells and data points, and the calculations occur in the background. Moving the modelling from Excel to R exposes all model components (and potential errors) and makes model redesign a more efficient process. (3) Replicability of analyses and model testing is made easier and more efficient. Each modification to the model during the redesign process requires tests for fidelity and reliability. The R environment saves time spent on model revision and allows more for time to be spent on rigorous validation of model component adaptations.

## 6.4.3 Capital Service Estimation

Direct capital service estimation is a feature of the model redesign that replaces the first iteration's 'capital service factor' approximation. A description of direct capital service estimation for Australian universities can also be found in (Moore et al., 2018a). The OECD (2001) provides a set of guidelines for estimating capital services under a variety of circumstances. The method adopted and advanced by the NRC for higher education capital service estimation uses figures for repairs and maintenance, rates for asset depreciation and

amortisation, as well as estimates for real rates of return (RRR). The NRC model benefited from external estimates of RRR on assets for universities in the United States (Sullivan et al., 2012). Without authoritative estimates of RRR for Australian university assets, a different approach is required.

Johnes (2006) demonstrated for UK universities that ‘interest payable’ (or interest bearing) can serve as an alternative proxy for capital service estimation. Interest payable is also a standard line item on Australian university financial reports. This can be combined with other standard line items on Australian university financial reports, including total depreciation and amortisation of assets and repairs and maintenance. This means that non-empirical approximations of depreciation rates and real rates of return to non-current assets found on university balance sheets is no longer necessary in the model redesign. This also serves to reduce the risk of applying rates in ways that deviate from standard Australian accounting practices.

While the NRC model took the line item of ‘repairs and maintenance’ and combined it with results from multiplying non-current assets figures against depreciation rates and real rates of return, the model redesign uses insight from both the NRC and Johnes (2006) to produce an Australian context specific estimate of capital services. No extra calculations are necessary. The model redesign takes the ‘repairs and maintenance’ and the ‘depreciation and amortisation’ figures from the annual expenses report, as well as the ‘current interest payable’ figures from the balance sheets. Together these three figures provide an empirical and theoretically sound proxy for capital services of Australian university assets that may vary per institution and vary each subsequent year. The more systematic, direct estimation of capital services in the model also increases confidence that capital input value shares are being computed more accurately.

#### **6.4.4 Input Smoothing**

While the first model iteration matched outputs with inputs from only the same calendar year, the redesigned model will account for a longer production process and subsequently link multiple years’ inputs to the outputs in the model. The need for doing so is also discussed in (Moore, Coates, et al., 2019). Higher education outputs inherently depend upon inputs from more than a single calendar year. Student degree programs and research projects generally take place over multiple years. A standard Törnqvist index considers the change between one year to the next from all data elements in the model. A more representative and reliable index for higher education would relate one year’s outputs to corresponding inputs used over the actual timespan that it took to produce the outputs, rather than drawing on figures from single annual reports.

The solution in model redesign is to match yearly outputs with the three consecutive years of inputs leading up to the output year. That is, for outputs from time  $t$ , corresponding input data comes from times  $t$ ,  $t - 1$ , and  $t - 2$ . The adjustment aims to increase the value and utility of measurement results, as well as increase the suitability of the model’s conceptual foundations for the higher education context.

The adjustment was made in response to findings from the model boundaries assessments from the formative evaluation. Evaluation findings raised concerns that specifications for the

first iteration of the model for the year-on-year Törnqvist indexes did not appropriately capture or represent the true production ‘technology’ or key work processes of universities. For the education function in the Australian context undergraduate course programs typically span three years. An undergraduate coursework completion could thus be considered as the product of three consecutive years of inputs. Postgraduate research programs are similar, ranging typically from two to four years (Australian Qualifications Framework Council, 2013). Research publications bear similarities. The time from submitting a research proposal, to proposal acceptance, research implementation, and then to publishing findings rarely takes place inside a single calendar year. It is an empirical question as to how long the research process takes at different institutions and in different fields of education, but results would no doubt be variable and inconsistent. This research uses a proxy. Australian Research Council grants under the National Competitive Grant Scheme typically run for two years or longer (Australian Research Council, 2015). The solution for model redesign is to associate outputs at time  $t$  with the arithmetic mean of inputs from times  $t$ ,  $t - 1$ , and  $t - 2$ . The solution is imperfect but addresses the issue raised in the evaluation above and will be open for scrutiny in the subsequent assessment.

Taking the arithmetic mean of three years of inputs is not arbitrary. First, there is precedent in the Australian system for determining public funding and financial research inputs for a given calendar year using weighted averages of performance data from two prior years (DOET, 2019). The rationale is that budgets are determined on an annual basis, but performance information from a single year is widely accepted as insufficient for portraying ongoing achievement. International work on performance measurement and performance funding warns against the risks of performance measures, mechanisms and targets that are too reactive to annual fluctuations (Liefner, 2003). There is a further rationale behind using the arithmetic mean across three years’ inputs, rather than some other type of average. Inputs for the education function may be considered as ‘front-loaded’, while inputs for research may be considered as ‘back-loaded’. That is, when an Australian university accepts an undergraduate student into a three-year program, it commits at the beginning of the program to the necessary inputs for facilitating the student’s education until completion. For research, from the time of proposal development until proposal acceptance, there is generally no commitment to inputs for the research project—and when a funding commitment is made, it is rare that all funds are given at once. Since the model measures joint productivity of education and research, the logic is that taking an arithmetic mean across three years for all inputs can balance the different funding natures of the two functions.

Both practically and conceptually, input smoothing suits the higher education context more so than does the standard industry, year-on-year Törnqvist index advanced by the OECD and NRC. The NRC issues a caveat about the risk of unreliable results for individual institutions because of the variability of data (Sullivan et al., 2012). Model input smoothing seeks to better represent institutional key work processes, so that the tool has increased utility for individual institutional contexts by reducing potential for artificial spikes and rebounds in measured productivity change.

#### 6.4.5 Output Value Weight Ranges

The redesigned model includes a function that runs the model multiple times for a range of values for the research and education component output index value weights,  $v_R$  and  $v_E$ . The adjustment speaks to multiple findings from the formative evaluation above. First, it addresses to the revelation that APO country participants consistently found value in measuring research and education productivity separately than how they chose to perform the analysis in their reports. Output value weight ranges provide a more technically and conceptually sound way to examine research and education productivity separately. Second, the new feature improves insight into unpacking final  $P_{\Delta t}$  metrics and helps promote understanding of what is driving measurement results. The value weight ranges emphasise or de-emphasise certain trends in the calculation, which can help pinpoint the drivers. Finally, output value weight ranges can provide utility for institutional decision-makers about how institutional performance may be matching with priorities. They provide more explicit information into what areas of performance are seeing the most productivity gains and allow for judgements to be made about whether performance trends match with institutional strategy and intentions.

The value weight ranges address one of the model's most significant empirical shortcomings. Findings from the model evaluation emphasised that the equitable designations of  $v_R = 0.5$  and  $v_E = 0.5$  follow system-wide academic ideals, but they likely do not accurately represent each institution's education and research priorities. Definitive designation of value weights for university outputs will likely not be possible unless all stakeholder interests in higher education outputs can be aligned. Further, any empirical method for determining the weights is not straight forward. Value weight assignment could be done based upon ranking strategic importance of key functions or work processes, surveying stakeholder interests, or gathering results from parallel quality and impact evaluations. The lack of a clear path forward for empirical determination means that the designation should be made based upon value judgements. Exposing options and alternatives in measured results based on ranges of value weights can provide a basis for conversation about what value judgements on different output types are most appropriate for individual institutions in context.

Value weight range accommodation in the model allows for separate analysis of education and research productivity without having to make non-empirical assumptions about how resources are allocated to the two academic functions. Little data is available on Australian universities to determine how resources are allocated to research and education. Further, the measurement model and its redesign are specified to estimate joint productivity of the two types of academic outputs. If more data were available on resource allocation to the separate academic functions, estimates of separate productivity could be made. Making non-empirical judgements, however, about resource allocation practices on the input-side of the model could lead to objectively incorrect results if better information on resource allocation emerged. The appropriate modelling choice is to test ranges for the research and education output value weights,  $v_R$  and  $v_E$ —which are subjective by their nature—to make inferences about research and education performance. The effects of testing the weights do not produce separate education and research productivity change estimates. Rather, they produce different joint productivity estimates that emphasise either research or education output trends.

Developing more appropriate methods to assign value weights in higher education productivity measurement has been addressed in literature (Massy, 2012), but no agreeable alternatives to common techniques like DEA have yet to emerge. Value weight range accommodation represents an improvement to common techniques. For example, the DEA techniques used in prior studies on Australian higher education productivity assign value weights based on computer algorithms. The algorithms do solve two problems. First, they allow researchers to avoid incorporating their own personal biases in an analysis. Second, outside stakeholders cannot claim that DEA representations serve to disproportionately benefit some institutions over others because DEA assigned value weights are intentionally designated to give each institution its most favourable productivity portrayal based on the data available. The issue with DEA, however, is that when considering contextual information and factors not represented in the dataset, algorithmic determination of value weights is likely more arbitrary than using researcher judgment guided by context. In addition, most DEA studies do not report the computer-assigned value weights. The weights remain shadow variables of the DEA calculation and cannot be scrutinised or validated.

#### 6.4.6 Cumulative Productivity Change

Cumulative productivity change calculation from  $P_{\Delta t}$  values represents a key innovation to the model that provides value and utility for interpreting measurement results. The adaptation to the model speaks to multiple findings from the evaluation of the model. A benefit of the DEA methods used in prior studies on Australian higher education is that the total change in productivity over a period with multiple time points can be tracked. The model showed only incremental changes from one time point to another without offering a wholistic interpretation of results over the entire period. The second model iteration allows for the total (or cumulative) change over the period to be tracked.

Cumulative productivity change is a derivative calculation from  $P_{\Delta t}$  values for clarifying the results of measurement. The shorthand for a cumulative productivity change index may be given by  $\Pi P_{\Delta t}$ . It tracks the culmination of productivity change indexes to signify total change over a period. A precise definition of  $\Pi P_{\Delta t}$  is given in **Appendix B**.

The benefit of calculating cumulative  $\Pi P_{\Delta t}$  indexes at each interval within the period is that they illustrate a more interpretable trend than what is shown with  $P_{\Delta t}$  indexes. The cumulative indexes give snowballing values for each yearly interval within the period. The result accumulates to a final productivity change value for the period equal the average or total change over the period. The calculations are best illustrated through example. The example that follows derives from results from the evaluation above.

One key finding from the Asia-Pacific evaluation is that an error in  $P_{\Delta t}$  results from the Pakistan APO country report was not identified in their analysis (Coates, 2017a). A cumulative productivity change calculation might have signalled the error and prompted revisions before reporting. Table 23 shows the corresponding  $\Pi P_{\Delta t}$  calculations that result from Pakistan's reported  $P_{\Delta t}$  indexes. The  $\Pi P_{\Delta t}$  calculations make the error more apparent. They signal a productivity increase by a dubious factor of greater than 6000 by the end of the five-year period. When reporting only year-on-year  $P_{\Delta t}$  results, the accumulation of incremental change over multiple time intervals can be lost if results are not correctly interpreted.

**Table 23: Pakistan reported productivity change indexes and corresponding  $\Pi P_{\Delta t}$  values**

| Year               | 2010 | 2011  | 2012   | 2013    | 2014    |
|--------------------|------|-------|--------|---------|---------|
| $P_{\Delta t}$     | 3.92 | 3.84  | 16.52  | 5.37    | 4.75    |
| $\Pi P_{\Delta t}$ | 3.92 | 15.05 | 248.67 | 1335.37 | 6343.01 |

#### 6.4.7 Productivity Threshold Estimation

Productivity threshold estimation represents a way to synthesise results from  $\Pi P_{\Delta t}$  index calculations across value weight ranges for  $v_R$  and  $v_E$ . This model adaptation is also discussed in Moore et al. (2018a). Under different output value weight specifications, cumulative productivity change trajectories can shift from more to less positive or vice versa. The utility of productivity threshold estimation is to illustrate upper and lower bounds of institutional productivity based on alternative output value judgements. The different trends that can result from  $\Pi P_{\Delta t}$  indexes under different output value judgements illustrate the range of plausible productivity change portrayals over a period from available data. Rather than settling on any single authoritative output value weighting regime, threshold estimation illustrates the range of portrayals that could result from the data. The analysis is made clear in the results from the model redesign that follow.

### 6.5 Model Re-design Results

The measurement results from the second iteration of the model on Australian data represent intermediate quantitative findings for this research project. They provide estimates for 38 universities and demonstrate the expanding capabilities of the measurement model. The prime objective of recording these intermediate findings is for use in the next phase of research. The expanded set of results illustrates the new scope for insight of the model and can demonstrate further potential value for institutional performance assessment to Australian stakeholders and experts. At this stage, no deep exploration of results is necessary, nor is any determination of findings implications. Once experts and stakeholders provide further feedback on the model, measurements can be explored more thoroughly. Presently, descriptions of major results are provided in the paragraphs that follow.

Table 24 shows productivity change measures for 38 universities using equal value weights on education and research. It shows the interpretive value of tracking cumulative productivity change over the period examined. The  $P_{\Delta t}$  figures show year-on-year per cent change estimates. The  $\Pi P_{\Delta t}$  figures show the total, or cumulative change since the beginning of the period. The 2013 cumulative change figures communicate the factor by which productivity has changed over the entire period. Australian Catholic University, for example, is measured at a level of productivity in 2013 that is 88.9% of what it was in 2007. Australian National University has a measured increase in productivity of 0.7% in 2013 over its 2007 levels.

**Table 24: Redesigned model productivity change and cumulative change results**

| <b>Institution</b>                  | <b>Metric</b>      | <b>2007</b> | <b>2008</b> | <b>2009</b> | <b>2010</b> | <b>2011</b> | <b>2012</b> | <b>2013</b> |
|-------------------------------------|--------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Australian Catholic University      | $P_{\Delta t}$     | 1           | 0.959       | 1.019       | 1.075       | 0.989       | 0.808       | 1.060       |
|                                     | $\Pi P_{\Delta t}$ | 1           | 0.959       | 0.977       | 1.050       | 1.039       | 0.839       | 0.889       |
| Australian National University      | $P_{\Delta t}$     | 1           | 1.025       | 0.968       | 1.080       | 0.970       | 0.963       | 1.006       |
|                                     | $\Pi P_{\Delta t}$ | 1           | 1.025       | 0.993       | 1.072       | 1.040       | 1.001       | 1.007       |
| Central Queensland University       | $P_{\Delta t}$     | 1           | 0.922       | 1.022       | 1.052       | 1.014       | 0.915       | 1.093       |
|                                     | $\Pi P_{\Delta t}$ | 1           | 0.922       | 0.943       | 0.991       | 1.006       | 0.920       | 1.005       |
| Charles Darwin University           | $P_{\Delta t}$     | 1           | 0.890       | 0.920       | 0.973       | 1.054       | 0.979       | 0.919       |
|                                     | $\Pi P_{\Delta t}$ | 1           | 0.890       | 0.819       | 0.797       | 0.840       | 0.822       | 0.755       |
| Charles Sturt University            | $P_{\Delta t}$     | 1           | 0.942       | 0.782       | 1.151       | 0.936       | 0.886       | 0.988       |
|                                     | $\Pi P_{\Delta t}$ | 1           | 0.942       | 0.737       | 0.848       | 0.794       | 0.703       | 0.695       |
| Curtin University of Technology     | $P_{\Delta t}$     | 1           | 0.903       | 0.953       | 0.941       | 0.963       | 0.954       | 0.976       |
|                                     | $\Pi P_{\Delta t}$ | 1           | 0.903       | 0.861       | 0.810       | 0.780       | 0.744       | 0.726       |
| Deakin University                   | $P_{\Delta t}$     | 1           | 1.025       | 0.982       | 0.917       | 0.942       | 1.047       | 0.961       |
|                                     | $\Pi P_{\Delta t}$ | 1           | 1.025       | 1.007       | 0.923       | 0.870       | 0.911       | 0.875       |
| Edith Cowan University              | $P_{\Delta t}$     | 1           | 1.053       | 0.934       | 0.989       | 0.957       | 0.935       | 1.025       |
|                                     | $\Pi P_{\Delta t}$ | 1           | 1.053       | 0.983       | 0.972       | 0.931       | 0.870       | 0.892       |
| Federation University Australia     | $P_{\Delta t}$     | 1           | 0.838       | 0.927       | 0.888       | 0.963       | 1.044       | 0.904       |
|                                     | $\Pi P_{\Delta t}$ | 1           | 0.838       | 0.777       | 0.691       | 0.665       | 0.694       | 0.627       |
| Flinders University                 | $P_{\Delta t}$     | 1           | 1.004       | 0.896       | 0.983       | 0.986       | 0.970       | 0.961       |
|                                     | $\Pi P_{\Delta t}$ | 1           | 1.004       | 0.900       | 0.884       | 0.872       | 0.846       | 0.813       |
| Griffith University                 | $P_{\Delta t}$     | 1           | 0.955       | 1.012       | 0.987       | 0.977       | 0.918       | 0.972       |
|                                     | $\Pi P_{\Delta t}$ | 1           | 0.955       | 0.967       | 0.955       | 0.933       | 0.856       | 0.833       |
| James Cook University               | $P_{\Delta t}$     | 1           | 0.911       | 1.067       | 0.933       | 1.036       | 0.937       | 0.965       |
|                                     | $\Pi P_{\Delta t}$ | 1           | 0.911       | 0.971       | 0.907       | 0.939       | 0.880       | 0.850       |
| La Trobe University                 | $P_{\Delta t}$     | 1           | 0.854       | 1.051       | 1.006       | 1.013       | 0.960       | 1.021       |
|                                     | $\Pi P_{\Delta t}$ | 1           | 0.854       | 0.897       | 0.903       | 0.914       | 0.877       | 0.896       |
| Macquarie University                | $P_{\Delta t}$     | 1           | 0.978       | 0.958       | 0.950       | 1.027       | 0.954       | 0.961       |
|                                     | $\Pi P_{\Delta t}$ | 1           | 0.978       | 0.937       | 0.890       | 0.914       | 0.872       | 0.838       |
| Monash University                   | $P_{\Delta t}$     | 1           | 0.948       | 0.983       | 0.960       | 0.984       | 0.953       | 1.021       |
|                                     | $\Pi P_{\Delta t}$ | 1           | 0.948       | 0.933       | 0.895       | 0.881       | 0.839       | 0.857       |
| Murdoch University                  | $P_{\Delta t}$     | 1           | 0.862       | 1.142       | 0.958       | 0.984       | 1.020       | 0.982       |
|                                     | $\Pi P_{\Delta t}$ | 1           | 0.862       | 0.984       | 0.943       | 0.928       | 0.946       | 0.929       |
| Queensland University of Technology | $P_{\Delta t}$     | 1           | 0.974       | 0.961       | 0.922       | 0.962       | 0.999       | 0.991       |
|                                     | $\Pi P_{\Delta t}$ | 1           | 0.974       | 0.936       | 0.863       | 0.830       | 0.829       | 0.821       |
| RMIT University                     | $P_{\Delta t}$     | 1           | 1.017       | 0.979       | 0.944       | 1.036       | 0.972       | 1.056       |
|                                     | $\Pi P_{\Delta t}$ | 1           | 1.017       | 0.996       | 0.940       | 0.974       | 0.947       | 0.999       |
| Southern Cross University           | $P_{\Delta t}$     | 1           | 0.960       | 0.940       | 0.874       | 0.930       | 0.998       | 0.964       |
|                                     | $\Pi P_{\Delta t}$ | 1           | 0.960       | 0.902       | 0.788       | 0.732       | 0.731       | 0.705       |
| Swinburne University of Technology  | $P_{\Delta t}$     | 1           | 1.064       | 0.969       | 1.063       | 0.948       | 0.997       | 1.041       |
|                                     | $\Pi P_{\Delta t}$ | 1           | 1.064       | 1.031       | 1.097       | 1.040       | 1.037       | 1.080       |
| The University of Melbourne         | $P_{\Delta t}$     | 1           | 0.958       | 0.940       | 0.929       | 1.017       | 0.965       | 0.973       |
|                                     | $\Pi P_{\Delta t}$ | 1           | 0.958       | 0.900       | 0.836       | 0.850       | 0.821       | 0.799       |
| The University of New England       | $P_{\Delta t}$     | 1           | 0.921       | 0.953       | 1.021       | 0.946       | 1.018       | 0.974       |
|                                     | $\Pi P_{\Delta t}$ | 1           | 0.921       | 0.877       | 0.895       | 0.847       | 0.862       | 0.839       |

|                                        |                    |   |       |       |       |       |       |       |
|----------------------------------------|--------------------|---|-------|-------|-------|-------|-------|-------|
| The University of New South Wales      | $P_{\Delta t}$     | 1 | 0.925 | 0.985 | 1.009 | 0.943 | 0.967 | 0.993 |
|                                        | $\Pi P_{\Delta t}$ | 1 | 0.925 | 0.911 | 0.920 | 0.867 | 0.839 | 0.833 |
| The University of Newcastle            | $P_{\Delta t}$     | 1 | 0.976 | 0.945 | 0.951 | 0.925 | 0.999 | 1.004 |
|                                        | $\Pi P_{\Delta t}$ | 1 | 0.976 | 0.922 | 0.877 | 0.811 | 0.810 | 0.814 |
| The University of Notre Dame Australia | $P_{\Delta t}$     | 1 | 1.089 | 0.773 | 1.194 | 0.727 | 0.950 | 1.257 |
|                                        | $\Pi P_{\Delta t}$ | 1 | 1.089 | 0.842 | 1.005 | 0.731 | 0.694 | 0.873 |
| The University of Queensland           | $P_{\Delta t}$     | 1 | 0.945 | 0.968 | 0.965 | 0.952 | 0.967 | 1.019 |
|                                        | $\Pi P_{\Delta t}$ | 1 | 0.945 | 0.915 | 0.883 | 0.841 | 0.813 | 0.828 |
| The University of Sydney               | $P_{\Delta t}$     | 1 | 0.955 | 0.910 | 0.986 | 0.977 | 0.966 | 0.991 |
|                                        | $\Pi P_{\Delta t}$ | 1 | 0.955 | 0.869 | 0.857 | 0.837 | 0.809 | 0.802 |
| The University of Western Australia    | $P_{\Delta t}$     | 1 | 0.955 | 0.970 | 0.990 | 1.025 | 1.002 | 0.943 |
|                                        | $\Pi P_{\Delta t}$ | 1 | 0.955 | 0.926 | 0.917 | 0.940 | 0.942 | 0.888 |
| University of Adelaide                 | $P_{\Delta t}$     | 1 | 0.981 | 0.933 | 0.979 | 0.928 | 0.967 | 0.976 |
|                                        | $\Pi P_{\Delta t}$ | 1 | 0.981 | 0.915 | 0.895 | 0.831 | 0.804 | 0.785 |
| University of Canberra                 | $P_{\Delta t}$     | 1 | 0.987 | 0.896 | 1.133 | 0.994 | 0.947 | 0.929 |
|                                        | $\Pi P_{\Delta t}$ | 1 | 0.987 | 0.885 | 1.003 | 0.996 | 0.943 | 0.876 |
| University of South Australia          | $P_{\Delta t}$     | 1 | 0.978 | 0.983 | 1.015 | 0.948 | 0.927 | 0.999 |
|                                        | $\Pi P_{\Delta t}$ | 1 | 0.978 | 0.961 | 0.975 | 0.925 | 0.857 | 0.857 |
| University of Southern Queensland      | $P_{\Delta t}$     | 1 | 0.895 | 1.047 | 0.926 | 0.967 | 1.098 | 0.903 |
|                                        | $\Pi P_{\Delta t}$ | 1 | 0.895 | 0.938 | 0.869 | 0.840 | 0.922 | 0.833 |
| University of Tasmania                 | $P_{\Delta t}$     | 1 | 0.940 | 0.950 | 0.986 | 0.954 | 1.014 | 0.975 |
|                                        | $\Pi P_{\Delta t}$ | 1 | 0.940 | 0.893 | 0.881 | 0.840 | 0.852 | 0.831 |
| University of Technology, Sydney       | $P_{\Delta t}$     | 1 | 0.945 | 0.980 | 0.948 | 0.949 | 0.944 | 1.028 |
|                                        | $\Pi P_{\Delta t}$ | 1 | 0.945 | 0.926 | 0.878 | 0.833 | 0.787 | 0.809 |
| University of the Sunshine Coast       | $P_{\Delta t}$     | 1 | 0.967 | 1.099 | 0.903 | 0.907 | 1.084 | 0.973 |
|                                        | $\Pi P_{\Delta t}$ | 1 | 0.967 | 1.063 | 0.960 | 0.870 | 0.944 | 0.919 |
| University of Western Sydney           | $P_{\Delta t}$     | 1 | 0.914 | 0.943 | 0.959 | 0.971 | 0.948 | 1.006 |
|                                        | $\Pi P_{\Delta t}$ | 1 | 0.914 | 0.862 | 0.827 | 0.803 | 0.761 | 0.765 |
| University of Wollongong               | $P_{\Delta t}$     | 1 | 0.965 | 1.019 | 0.921 | 0.980 | 0.940 | 1.004 |
|                                        | $\Pi P_{\Delta t}$ | 1 | 0.965 | 0.983 | 0.905 | 0.887 | 0.834 | 0.837 |
| Victoria University                    | $P_{\Delta t}$     | 1 | 1.081 | 0.978 | 0.991 | 0.938 | 0.971 | 0.954 |
|                                        | $\Pi P_{\Delta t}$ | 1 | 1.081 | 1.057 | 1.048 | 0.983 | 0.954 | 0.910 |

Table 25 provides a glance at the differences in productivity estimates between the first and second iterations of the model in this research and with respect to estimates given by Moradi-Motlagh et al. (2016) from their DEA model. While there are differences in measured results and relative rankings between the initial model in this research and its re-design, the biggest discrepancies can be seen between both models tested in this research and the DEA model. The DEA model bears few similarities in terms of relative ranks between institutions, and provides much more positive estimates of change over the period.

**Table 25: Cumulative change comparison for different models 2007-2013**

| Institution                        | Initial Model |                    | Redesigned Model |                    | *Model DEA MJH |                    |
|------------------------------------|---------------|--------------------|------------------|--------------------|----------------|--------------------|
|                                    | Rank          | $\Pi P_{\Delta t}$ | Rank             | $\Pi P_{\Delta t}$ | Rank           | $\Pi P_{\Delta t}$ |
| Swinburne University of Technology | 1             | 1.080              | 1                | 1.156              | 17             | 1.172              |
| Australian National University     | 2             | 1.007              | 3                | 1.010              | 10             | 1.228              |

|                                        |    |       |    |       |    |       |
|----------------------------------------|----|-------|----|-------|----|-------|
| Central Queensland University          | 3  | 1.005 | 2  | 1.040 | 25 | 1.116 |
| RMIT University                        | 4  | 0.999 | 5  | 0.988 | 9  | 1.252 |
| Murdoch University                     | 5  | 0.929 | 7  | 0.961 | 1  | 1.513 |
| University of the Sunshine Coast       | 6  | 0.919 | 6  | 0.974 | 19 | 1.147 |
| Victoria University                    | 7  | 0.910 | 4  | 1.006 | 11 | 1.224 |
| La Trobe University                    | 8  | 0.896 | 9  | 0.944 | 13 | 1.219 |
| Edith Cowan University                 | 9  | 0.892 | 12 | 0.899 | 21 | 1.127 |
| Australian Catholic University         | 10 | 0.889 | 19 | 0.859 | 24 | 1.116 |
| The University of Western Australia    | 11 | 0.888 | 8  | 0.948 | 3  | 1.396 |
| University of Canberra                 | 12 | 0.876 | 13 | 0.898 | 4  | 1.325 |
| Deakin University                      | 13 | 0.875 | 25 | 0.840 | 16 | 1.181 |
| The University of Notre Dame Australia | 14 | 0.873 | 24 | 0.843 | NA | NA    |
| Monash University                      | 15 | 0.857 | 11 | 0.904 | 7  | 1.254 |
| University of South Australia          | 16 | 0.857 | 15 | 0.885 | 15 | 1.184 |
| James Cook University                  | 17 | 0.850 | 20 | 0.856 | 29 | 1.068 |
| The University of New England          | 18 | 0.839 | 32 | 0.788 | 14 | 1.192 |
| Macquarie University                   | 19 | 0.838 | 29 | 0.812 | 30 | 1.066 |
| University of Wollongong               | 20 | 0.837 | 21 | 0.850 | 32 | 1.041 |
| The University of New South Wales      | 21 | 0.833 | 16 | 0.880 | 2  | 1.43  |
| Griffith University                    | 22 | 0.833 | 28 | 0.824 | 8  | 1.252 |
| University of Southern Queensland      | 23 | 0.833 | 17 | 0.863 | 27 | 1.094 |
| University of Tasmania                 | 24 | 0.831 | 10 | 0.926 | 12 | 1.22  |
| The University of Queensland           | 25 | 0.828 | 14 | 0.890 | 18 | 1.17  |
| Queensland University of Technology    | 26 | 0.821 | 27 | 0.833 | 28 | 1.076 |
| The University of Newcastle            | 27 | 0.814 | 22 | 0.847 | 20 | 1.144 |
| Flinders University                    | 28 | 0.813 | 31 | 0.795 | 22 | 1.123 |
| University of Technology, Sydney       | 29 | 0.809 | 26 | 0.834 | 26 | 1.1   |
| The University of Sydney               | 30 | 0.802 | 18 | 0.862 | 6  | 1.258 |
| The University of Melbourne            | 31 | 0.799 | 23 | 0.844 | 5  | 1.28  |
| University of Adelaide                 | 32 | 0.785 | 30 | 0.804 | 34 | 0.989 |
| University of Western Sydney           | 33 | 0.765 | 33 | 0.761 | 23 | 1.119 |
| Charles Darwin University              | 34 | 0.755 | 34 | 0.749 | 35 | 0.988 |
| Curtin University of Technology        | 35 | 0.726 | 35 | 0.739 | 33 | 0.997 |
| Southern Cross University              | 36 | 0.705 | 36 | 0.738 | 37 | 0.957 |
| Charles Sturt University               | 37 | 0.695 | 37 | 0.726 | 36 | 0.957 |
| Federation University Australia        | 38 | 0.627 | 38 | 0.722 | 31 | 1.062 |

\*Model DEA MJH data comes from Moradi-Motlagh, et al. (2016).

Table 26 shows the effect on measured results that different emphasis on data elements can have on the field of institutions. The table provides five different rankings of university productivity change over the period with respect to different weightings of education and research outputs. Among 38 institutions, a group of 11 universities can climb or fall in the ranks by 15 places, depending on how different data elements are emphasised during calculation. The results speak to the sensitivity of higher education data in Australia to different aggregation methods.

**Table 26: Redesigned model rankings for different value weight schemes 2007-2013**

| <b>Institution</b>                     | <b>Rank</b><br>$v_E / v_R$<br><b>0.9 / 0.1</b> | <b>Rank</b><br>$v_E / v_R$<br><b>0.7 / 0.3</b> | <b>Rank</b><br>$v_E / v_R$<br><b>0.5 / 0.5</b> | <b>Rank</b><br>$v_E / v_R$<br><b>0.3 / 0.7</b> | <b>Rank</b><br>$v_E / v_R$<br><b>0.1 / 0.9</b> | <b>Potential<br/>Climb in<br/>Rank</b> |
|----------------------------------------|------------------------------------------------|------------------------------------------------|------------------------------------------------|------------------------------------------------|------------------------------------------------|----------------------------------------|
| Murdoch University                     | 1                                              | 2                                              | 5                                              | 26                                             | 34                                             | ^33                                    |
| Central Queensland University          | 31                                             | 9                                              | 3                                              | 1                                              | 1                                              | ^30                                    |
| Charles Darwin University              | 7                                              | 18                                             | 34                                             | 37                                             | 37                                             | ^30                                    |
| University of Southern Queensland      | 35                                             | 33                                             | 23                                             | 11                                             | 6                                              | ^29                                    |
| University of South Australia          | 32                                             | 28                                             | 16                                             | 8                                              | 5                                              | ^27                                    |
| University of the Sunshine Coast       | 25                                             | 12                                             | 6                                              | 4                                              | 3                                              | ^22                                    |
| Queensland University of Technology    | 33                                             | 32                                             | 26                                             | 17                                             | 12                                             | ^21                                    |
| The University of New England          | 9                                              | 13                                             | 18                                             | 27                                             | 27                                             | ^18                                    |
| Flinders University                    | 13                                             | 21                                             | 28                                             | 30                                             | 29                                             | ^17                                    |
| The University of Melbourne            | 16                                             | 27                                             | 31                                             | 31                                             | 30                                             | ^15                                    |
| The University of Western Australia    | 6                                              | 6                                              | 11                                             | 15                                             | 21                                             | ^15                                    |
| Australian Catholic University         | 19                                             | 14                                             | 10                                             | 6                                              | 7                                              | ^13                                    |
| James Cook University                  | 11                                             | 15                                             | 17                                             | 21                                             | 24                                             | ^13                                    |
| University of Canberra                 | 8                                              | 7                                              | 12                                             | 16                                             | 19                                             | ^12                                    |
| Griffith University                    | 27                                             | 26                                             | 22                                             | 18                                             | 16                                             | ^11                                    |
| The University of Newcastle            | 18                                             | 25                                             | 27                                             | 29                                             | 28                                             | ^11                                    |
| The University of Notre Dame Australia | 21                                             | 16                                             | 14                                             | 10                                             | 10                                             | ^11                                    |
| University of Technology, Sydney       | 29                                             | 31                                             | 29                                             | 25                                             | 20                                             | ^11                                    |
| University of Tasmania                 | 15                                             | 19                                             | 24                                             | 24                                             | 25                                             | ^10                                    |
| Victoria University                    | 5                                              | 5                                              | 7                                              | 12                                             | 15                                             | ^10                                    |
| Edith Cowan University                 | 14                                             | 10                                             | 9                                              | 7                                              | 8                                              | ^7                                     |
| Southern Cross University              | 38                                             | 37                                             | 36                                             | 34                                             | 31                                             | ^7                                     |
| RMIT University                        | 3                                              | 3                                              | 4                                              | 5                                              | 9                                              | ^6                                     |
| University of Adelaide                 | 26                                             | 29                                             | 32                                             | 32                                             | 32                                             | ^6                                     |
| Macquarie University                   | 22                                             | 22                                             | 19                                             | 19                                             | 17                                             | ^5                                     |
| University of Wollongong               | 23                                             | 23                                             | 20                                             | 20                                             | 18                                             | ^5                                     |
| The University of Sydney               | 28                                             | 30                                             | 30                                             | 28                                             | 26                                             | ^4                                     |
| University of Western Sydney           | 30                                             | 34                                             | 33                                             | 33                                             | 33                                             | ^4                                     |
| La Trobe University                    | 10                                             | 8                                              | 8                                              | 9                                              | 11                                             | ^3                                     |
| Monash University                      | 17                                             | 17                                             | 15                                             | 14                                             | 14                                             | ^3                                     |
| The University of New South Wales      | 20                                             | 20                                             | 21                                             | 22                                             | 23                                             | ^3                                     |
| The University of Queensland           | 24                                             | 24                                             | 25                                             | 23                                             | 22                                             | ^3                                     |
| Australian National University         | 4                                              | 4                                              | 2                                              | 3                                              | 4                                              | ^2                                     |
| Deakin University                      | 12                                             | 11                                             | 13                                             | 13                                             | 13                                             | ^2                                     |
| Charles Sturt University               | 36                                             | 36                                             | 37                                             | 36                                             | 36                                             | ^1                                     |
| Curtin University of Technology        | 34                                             | 35                                             | 35                                             | 35                                             | 35                                             | ^1                                     |
| Federation University Ballarat         | 37                                             | 38                                             | 38                                             | 38                                             | 38                                             | ^1                                     |
| Swinburne University of Technology     | 2                                              | 1                                              | 1                                              | 2                                              | 2                                              | ^1                                     |

Figure 21 shows how cumulative productivity change estimation in conjunction with different output value weights can provide a range of plausible productivity estimates for different institutions. The value weights range from a 70% emphasis on education and 30%

emphasis on research, to equal value weights, to a 30% emphasis on education and a 70% emphasis on research. Results show the extent to which individual institutions are affected by value weighting schemes in the data aggregation method. Moore et al. (2018a) explains that larger thresholds are neither ‘good’ nor ‘bad’, but they signal a more complex question about how well an institution is performing in terms of its inputs and outputs, as well as whether performance aligns with policies and priorities. Consider Murdoch University and Central Queensland University from Table 26. These institutions tell opposite stories of performance when emphasis switches from education to research. When emphasis is equal, however, their productivity change appears similar. Findings highlight the importance of entertaining multiple scenarios, and they provide targeted inroads for conducting supplementary analysis and highlight the importance of exercising judgement and incorporating contextual knowledge when determining whether measured trends match strategic intentions. Figure 21 provides a visualisation of the extent to which alternative scenarios are important to consider for each institution.

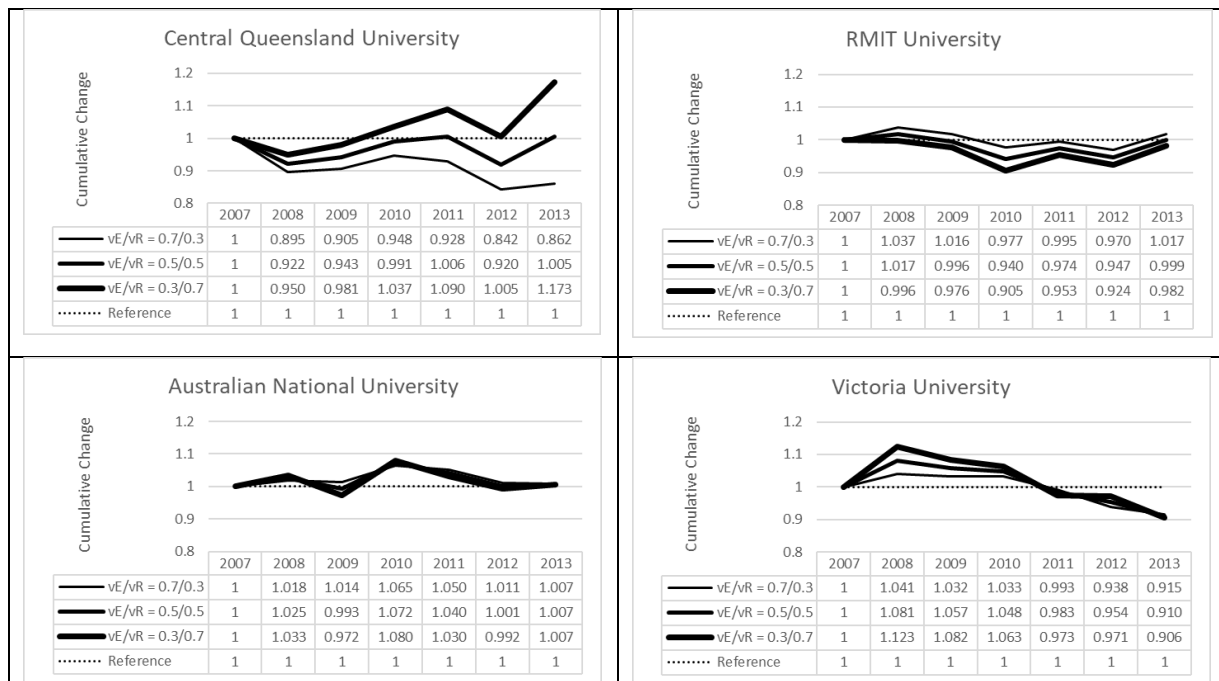


Figure 21:  $PP_{\Delta t}$  thresholds for four universities

## 6.6 Summary

### 6.6.1 Focus and Contributions

This chapter evaluated evidence from ten counties’ application of the productivity change measurement model under design in this research. Measurement results and experiences from nine APO countries and from Australia were examined using design science research guidelines. The design science model evaluation criteria guided investigation of model

specifications, model boundaries and scope for insight, as well as the model's potential for generating novel and useful information that aligns with countries' performance interests. The chapter also demonstrated numerous practical considerations for executing a university productivity measurement initiative.

A central contribution of this chapter has been exposure of the inherent variability and complexity associated with higher education productivity data and with the overall exercise of measurement. It serves as a first-hand demonstration of the need for new approaches to measurement. The contributions of this chapter represent significant, evidence-based improvements to standard practice for higher education productivity measurement using TI methods. While the chapter concludes with technical adjustments to standard TI specifications, the work in this chapter differs from comparable quantitative studies on higher education productivity measurement. Adaptations to the model in this research do not represent statistical adjustments for generating more precise and robust estimates in relation to a dataset. Rather, they represent experience-based improvements for better capturing the unique nature of higher education productive processes. Model adaptations in this chapter concern the selection and treatment of specific types of data elements. They provide solutions for embracing known variability between and within institutions—rather than creating decision-rules to settle on a single productivity portrayal deemed most appropriate by a computer algorithm. Finally, the adjustments represent a paradigm shift in higher education productivity measurement practice. Rather than making incremental adjustments to frontier estimation techniques most commonly used in the scholarly literature, these model adaptations move measurement practice in a different direction. They serve to capture authentic institutional performance trends and to represent the interests of stakeholders who intend to use results for decision-making.

### 6.6.2 Considerations and Limitations Arising

This chapter demonstrated the inherent imprecision and variability associated with higher education productivity measurement, but also it served to characterise and parameterise the complexity. In moving forward, the remainder of measurement and assessment efforts focus on Australian universities. Important to consider is the extent to which outliers and anomalies in collected data—and unobserved contextual phenomena—may impact measured results. A exhibited in Figure 21:  $II\bar{P}_{\Delta t}$  thresholds for four universities, the new 'productivity threshold' feature of the model addresses this reality. Institutional productivity estimates from this point forward are no longer represented as single, narrow trends. The measurement objective in subsequent analyses is to present plausible upper and lower bounds on institutional productivity portrayals, representative of the complexity of individual institutions' data.

The layered, aggregate nature of source data for productivity estimates means that any institution-level measurement tool will be associated with reliability issues. No measurement portrayal of university productivity should be mistaken as a definitive. Reliability issues are explicated in Table 25: Cumulative change comparison for different models 2007-2013. The productivity thresholds, or intervals, estimated from this point forward may be representative of statistical uncertainty, but they are not confidence intervals based on statistical error. Rather, the productivity thresholds presented in subsequent analyses represent real alternative

interpretations of productivity based on the data elements used in calculation. The upper and lower bounds have inherent meanings and present alternative perspectives on performance based on different emphases and interests in research and education outputs.

## CHAPTER 7: FORMAL EVALUATION AND MODEL ENHANCEMENT

*There's a fundamental mismatch between what we expect universities to do and what they can be measured as doing.*

–Study Participant KS.4  
Higher Education Policy Advisor

*With your model, you can delve more into what's causing the trends, and that's what's important—the 'why'.*

–Study Participant KS.3  
University Former Faculty Dean

### **Abstract:**

This chapter describes the final evaluation of the model in this research. The prime purpose of the chapter is to illustrate how contextual and practical concerns in higher education can be systematically addressed and built into the specifications of a measurement model. The evaluation incorporates interviews with a purposive sample of 20 participants, including ten key stakeholders and ten technical experts in Australia. Findings address research questions Q1, Q2 and Q3. First, it provides details on how Australian stakeholders view and understand higher education productivity. Second, feedback from interview participants is used to redesign and define final specifications for the productivity measurement model in this research. Third, participants provide diverse, experiential perspectives on the state and trends of Australian university productivity. Findings from the evaluation support proof-of-concept of the model and inform new adaptations to improve productivity estimates. Key adaptations include: expanding the model's input-output framework; adjusting estimates for currency inflation; tracking dynamics of individual data elements; and additional specifications for the comparison of joint and separate productivity, as well as SFP and TFP. Notable findings from this chapter are also shared in Moore (2018).

### **Key Words:**

Interviews, paradigms, policy, operations, model, adaptations

## 7.1 Introduction

This chapter presents findings from Stages 5 and 6 of the design research. Model evaluation is conducted through analysing response data from 20 interview participants selected to assess the productivity measurement model and its results. Participants also provided contextual and experiential perspectives on the state of higher education productivity in Australia. The body of evidence generated in the chapter addresses questions Q1, Q2 and Q3. The chapter is organised into three major sections: interview methods, expert evaluation and model enhancement. Results from the evaluation reveal specific strengths and limitations of the model. Interview results characterise diverse examples of how different stakeholders understand and conceptualise higher education productivity. Evaluation results also reveal interview participants' ideas for what productivity improvements could be made in Australian higher education and what constraints limit progress.



**Figure 22: Research Stages 5 & 6**

The purpose of this chapter is to provide the preponderance of evidence and rationale for all technical specifications that compose the final contextualised measurement model in this research. The product of the chapter is a fit-for-purpose measurement model for the Australian context. The objective of the expert and stakeholder evaluation is to allow for the incorporation of both technical and lay perspectives on the measurement model from individuals who have vested interests in its results. As with the previous chapter, the evaluation documented in Section 7.3 consistently follows Williamson and Johanson's (2013) four-point design science model assessment criteria. As described below in Section 7.2, however, rather than being based on quantitative measurement results from different international tests of the model, the evaluation in this chapter synthesises the qualitative responses of interview participants. In this way, the evaluation in this chapter is underpinned with grounded theory, guided by the questions posed to participants during interviews. The questions elicit critical assessments of the model and its results subsequent to their having reviewed a provided description of the model and sample findings on Australian universities.

In the interest of framing subsequent discussion, it is helpful to provide a description of some of the key findings and results from the chapter. First, participants were asked, not only to provide feedback on model components, but also on the overarching concept of productivity and its measurement in higher education. A synthesis of participant responses on how they understood higher education productivity uncovered four distinct operative paradigms for

conceptualising the topic. The paradigms are explained below and have been named: Value creation productivity, efficiency productivity, process productivity, and socio-technical productivity. Learnings from assessments of the model also culminated in six clear options for redesign. The six model adaptations are as follows:

1. Adjusting estimates for currency inflation and the real value of money
2. Tracking individual data element trends to associate with productivity estimates
3. Incorporating staff labour inputs to track single factor productivity
4. Calculating separate education and research productivity
5. Scaling productivity change trends with absolute productivity approximations
6. Restructuring research outputs for increased internal consistency

## **7.2 Evaluation Methods**

### **7.2.1 Methodology**

As with the previous chapter, the model evaluation in this chapter is guided by Williamson & Johanson's (2013) design science model assessment criteria. The evaluation in this chapter, however, considers the qualitative responses from interview participants who individually assessed the model based on their personal expertise and their stake in Australian higher education. Participants attributes, participants selection, interview question formation, and methods for data collection and analysis are described in sections 7.2.2 and 7.2.3.

The interview response analysis operates under the principal rationale of the design science research 'relevance cycle', as emphasised in Figure 23. The qualitative data analysis and evaluation represents the primary effort in this research to tie modelling and measurement decisions to the application domain where university productivity change measurement is set to take place—the Australian higher education system. The three design research cycles are concurrent, however, and the interview discussion guide was also designed with a minor objective to elicit general responses on university productivity and its measurement. Such responses are intended to add to the foundations and general knowledge base on higher education productivity measurement. Details are described in the methods below.

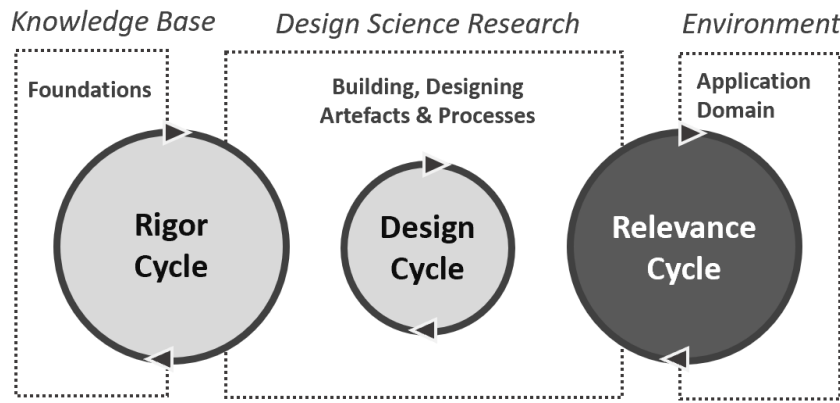


**Figure 23: Design science relevance cycle**

### 7.2.2 Participants

A purposive sample of 20 participants was selected to provide technical and contextual perspectives productivity and productivity measurement in Australian higher education. They were selected as key informants to aid in further development of the measurement model and to provide targeted insight for addressing all research questions. Participants were not selected randomly or as representative of a target population. Participants were selected, however, to provide diverse perspectives. Selection included participants who could represent key higher education interests with respect to Clark's (1983) Triangle of Coordination, the State, the Market and the Academy. With reference to Table 27, participants KS.4, TE.7 and TE.8 were selected specifically to represent state interests. Participants TE.2, TE.5 and TE.6 were selected to represent market interests. Participants KS.7, KS.9 and KS.10 were selected to represent academic interests. Most participants, however, had diverse experiences in multiple domains and were selected according to the following criteria.

Participants were selected principally as either (A) key stakeholders (KS) or (B) technical experts (TE). Key stakeholders were chosen based on their current or former roles as institutional leaders who use performance data to make decisions. They include university executives and deans, as well as advisors to executives. Two academic staff members were also interviewed as key stakeholders to provide perspectives from those who perform the key work processes ultimately being represented in productivity measurement. Technical experts were chosen based on their roles or experience in analysing productivity and efficiency data and in composing reports to be used for decision-making. They include consultants, researchers, officials of public agencies, and university professional staff. With respect to participants' primary full-time roles, they come from eight different Australian universities, four different consulting firms, and one public sector agency. Selection criteria did not include minimum years' experience, but rather focused on known public accomplishments and some snowball sampling using recommendations and referrals from other participants. Table 27 shows participant codes and gives monikers for different universities.

**Table 27: Participant Codes**

| Code | Academic Role                           | Public Sector or Professional Role |
|------|-----------------------------------------|------------------------------------|
| KS.1 | University A Advisor to Vice Chancellor | NA                                 |
| KS.2 | University B Chief Financial Officer    | NA                                 |

|              |                                               |                                                             |
|--------------|-----------------------------------------------|-------------------------------------------------------------|
| <b>KS.3</b>  | University C Faculty Dean (former)            | Project Lead, international higher education ranking system |
| <b>KS.4</b>  | NA                                            | Higher Education Policy Advisor, Small firm                 |
| <b>KS.5</b>  | University D Chief Operating Officer (former) | Financial Consultant, Small Firm                            |
| <b>KS.6</b>  | University A Vice Chancellor                  | NA                                                          |
| <b>KS.7</b>  | University C Academic Board President         | NA                                                          |
| <b>KS.8</b>  | University D Deputy Vice Chancellor           | NA                                                          |
| <b>KS.9</b>  | University C Academic Program Director        | NA                                                          |
| <b>KS.10</b> | University C Associate Professor              | NA                                                          |
| <b>TE.1</b>  | University E Lecturer                         | Private Sector Consultant, Private Practice                 |
| <b>TE.2</b>  | NA                                            | Education Consultant, Large Firm                            |
| <b>TE.3</b>  | University F Professor                        | Health Sector Economist                                     |
| <b>TE.4</b>  | University G Professor                        | NA                                                          |
| <b>TE.5</b>  | NA                                            | Higher Education Operations Consultant, Small Firm          |
| <b>TE.6</b>  | NA                                            | Higher Education Operations Consultant, Small Firm          |
| <b>TE.7</b>  | University E Professor                        | Public Policy Architect                                     |
| <b>TE.8</b>  | NA                                            | Public Sector Economist, Government Agency                  |
| <b>TE.9</b>  | University H Chief Financial Officer (former) | Higher Education Consultant, Private Practice               |
| <b>TE.10</b> | University G non-academic staff member        | NA                                                          |

It is important to reiterate that the group of participants is not a representative random sample. Key informants were selected to provide targeted feedback as experts and stakeholders to assist in evaluating the measurement model under development in this research. Accordingly, no thematic analysis is performed with respect to participant attributes or demographics. Analysis is performed only with respect to evaluation objectives and research questions. In practice this means that, even if patterns from within the sample emerge—for example—with respect to males or females, or with respect to academic and non-academic university staff, these patterns are not reported. There should be no mistaking that patterns among subgroups within this sample could have emerged by random chance because the subgroups are small.

Evidence generated across all interviews, however, should instil confidence that collected data are sufficient for informing important revisions to the measurement model, as well as to characterise productivity trends and understandings within the Australian system. Interviews were discontinued once responses reached saturation point. Participant feedback and responses to the interview questions became repetitive by the 15<sup>th</sup> interview. Five more interviews were conducted at this point to ensure saturation had been reached. As anticipated, little new information was exposed with the five final interviews, but they served an important purpose of cementing and emphasising certain key phenomena that had been raised previously.

### 7.2.3 Data Collection and Analysis

The first step of preparation leading to the interviews was to develop an artefact that represented key aspects of the measurement model and captured the types of results it could

produce. The intention was to send the artefact to interviewees prior to their interview, so that in order to avoid too much being spent on interpreting the model during the interview. Further, because the model has many complex elements, it was anticipated that conversation would be aided if participants could enter the interview with preliminary thoughts about the model. Two different artefacts were developed, one for the group of key stakeholders and the other for the group of technical experts. Each provides a summary of the model and its results, but the artefact for technical experts gives more detail about mathematical specifications. They can be seen in **Appendix C** and **Appendix D**.

The next step was to design an interview discussion guide to facilitate discussion during the interviews. It was decided that interviews with both participant groups should use the same interview form. The intention of the questions is for both groups to provide some degree of both technical and contextual feedback so their responses could be used to address both the design science model assessment criteria and the main research questions of this project. The discussion form was designed to allow for this while not using the direct technical language of the design science literature. The form was thus designed to solicit the types of information listed below, and a copy of the interview form can be found in **Appendix E**.

1. Views and ideas about the meaning term ‘productivity’ in the context of higher education
2. Initial reactions to the provided model summary
3. Ideas about what characteristics of the model could be changed or revised
4. Views on critical productivity issues in Australian higher education
5. Views on constraints and barriers to productivity improvement

The tool was developed in consultation with experts and was piloted with a former university executive. The final interview form remained unchanged across all 20 interviews. Participants’ consent was obtained before commencing and recording each interview. The discussions lasted for a duration of 45 to 75 minutes. Interviews were transcribed verbatim in Microsoft Word. Thematic analysis was performed on the transcripts to detect emergent patterns in the data with respect to Williamson & Johanson's (2013) design science model assessment criteria, and patterns were identified to help directly address the research questions in this project. Findings from the analysis of transcripts are discussed in the section below.

## **7.3 Expert and Stakeholder Evaluation**

### **7.3.1 Underlying Constructs Evaluation**

#### **7.3.1|a. *Productivity paradigms***

Findings from the constructs evaluation reveal four underlying paradigms for understanding productivity. The themes may be described as broad constructs or conceptualisations of productivity that are more abstract and more fundamental than those related to tangible model

specifications. The paradigms may be thought of as meta-constructs in relations to models for productivity measurement. Identifying the paradigms helps to address research question Q1.

The productivity paradigms may be understood as fundamental ideas held by interviewees for constructing a definition of higher education productivity. Identification of the paradigms emerged from thematic analysis of responses to the first question from the interview discussion guide. Although patterns in conceptualising productivity could be identified, articulating a concise set of productivity paradigms was not a straightforward task. As noted in Chapter 2, Tangen (2005) explains that, though the term is widely used, there is lack of common agreement about what it means, and precedents in the scholarly literature on conceptualising productivity had to be used in order to initiate thematic analysis. Three notable works guided analysis in terms of articulating distinct categories, Ghobadian and Husband (1990), Massy (2017) and Pritchard (1995). Each of the studies determines a unique taxonomy identifying underlying ideas about productivity. Table 29 provides concise definitions and explanations of the paradigms identified in this research. The paradigms are termed ‘value creation productivity’, ‘efficiency productivity’, ‘process-oriented productivity’, and ‘pragmatic productivity’. The following paragraphs further explain the paradigms and provide the empirical foundation for their identification with illustrative participant quotes.

Table 28 summarises three documented categorization schemes for productivity paradigms and allows for comparison between the taxonomies.

The current research uses the productivity paradigms already identified in the literature and considers interview participant responses to identify four new paradigms unique to the higher education context. It should be noted, however, that it is possible for individuals to use different productivity paradigms based on context. Thus, while the four productivity paradigms identified in this research represent mutually exclusive ideas, they are not exclusive to individuals. Rather a single participant may have exhibited a different framing of productivity depending on the topic being discussed. Further details about the paradigms being applied to different higher education topics is discussed in more detail in the next chapter. Table 29 provides concise definitions and explanations of the paradigms identified in this research. The paradigms are termed ‘value creation productivity’, ‘efficiency productivity’, ‘process-oriented productivity’, and ‘pragmatic productivity’. The following paragraphs further explain the paradigms and provide the empirical foundation for their identification with illustrative participant quotes.

**Table 28: Different classifications of productivity paradigms in recent literature**

|                                        |                                                       |                                                                        |                                                                                |
|----------------------------------------|-------------------------------------------------------|------------------------------------------------------------------------|--------------------------------------------------------------------------------|
| Ghobadian and Husband (1990) Paradigms | <i>Technological concept</i><br>(Output vs. input)    | <i>Engineering concept</i><br><br>(Actual output vs. potential output) | <i>Economists concept</i><br><br>(Resource and resource allocation efficiency) |
| Pritchard (1995) Paradigms             | <i>Techno-economic approach</i><br>(Output vs. input) | <i>Efficiency-effectiveness approach</i>                               | <i>Performance approach</i><br><br>(wide consideration of any functional       |

|                        |                                                                |                                                        |                                                  |
|------------------------|----------------------------------------------------------------|--------------------------------------------------------|--------------------------------------------------|
|                        |                                                                | (Output vs. input and output vs. goal)                 | improvement)                                     |
| Massy (2017) Paradigms | <i>Output paradigm</i><br>(More output = greater productivity) | <i>Marketplace paradigm</i><br>(Margin before mission) | <i>Economists paradigm</i><br>(Output vs. input) |

**Table 29: Productivity paradigms identified in the current research**

| <b>Paradigm</b>               | <b>Description</b>                                                                                                                                                                                                                                                                          |
|-------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Value Creation Productivity   | Higher productivity means <b>more output produced in less time</b> . There is no reconciliation with physical input. A productive institution is one that generates more valuable outputs in the allotted timeframe, regardless of what is physically consumed during productive processes. |
| Efficiency Productivity       | Higher productivity means <b>consuming less to achieve desired results</b> . Indicators are usually financial. A productive university exhibits desirable performance on balance sheets, income statements and budgets.                                                                     |
| Process-Oriented Productivity | Higher productivity is <b>not demonstrable through inputs and outputs</b> . A productive institution is one with high quality academic work, where the resultant <i>quality</i> of outputs supersedes indicators of quantity.                                                               |
| Pragmatic Productivity        | Higher productivity means <b>more efficient creation of more value</b> . A productive institution pursues strategic objectives by innovating and improving functions to maximise value under known physical constraints for the benefit of students and users/consumers of research.        |

### **7.3.1|b. Value creation productivity**

The value creation productivity paradigm for higher education operates under the reasoning that ‘more is more’. It is based on Massy’s (2017) ‘output paradigm’. While the value creation paradigm may not dominate mindset of any of the interview participants, this mode of thinking and framing of higher education productivity and performance was raised repeatedly. The framing arises commonly with respect to international rankings. Most institutional rankings systems emphasise indicators associated with intuitional outputs, outcomes, and reputation—rather than the costs associated with high levels of production and value creation. Participant KS.6 noted, “in a lot of the rankings, you move up just by being bigger”. Nevertheless, as claimed by KS.2, Australian institutions are “drawn into playing the rankings game”. In this case, the rankings game refers to maximising certain measured value indicators (quantity or quality) at indiscriminate cost. Two further participant quotes illustrate the influence of the value creation productivity paradigm in higher education:

*It's interesting because we always think about productivity—at least I do—in terms of research outputs.*

–KS.10

*In the name of investment and in the name of improving a university, people will pursue and prefer—in my experience—a quality agenda, even if it will cost more... at the expense of productivity, if you put a hard measure on it.*

–KS.4

### **7.3.1|c. Efficiency productivity**

The efficiency productivity paradigm for higher education is focused on inputs and largely financial. It may be thought of as a merging of Ghobadian and Husband's (1990) 'economists concept' and Massy's (2017) 'marketplace paradigm'. Participants who demonstrated this mode of thinking often described key university productivity indicators as factors related to operational margin and ratios of staff numbers to staff expenditures. The perspective is not blind, however, to academic work or the quality of academic outputs. The efficiency paradigm, rather, positions the framing of productivity around maximising operational margin, so that sufficient funds are available to invest back into the institution for quality improvements to academic programs.

*In the for-profit sector, you have access to equity because you can raise capital from investors. When you're a public university, you have no money to invest if you don't generate cash as reported on the income statement.*

–KS.5

*I think the best way to look at it is a cost vs. outcome perspective, in terms of what you're trying to achieve. From an education and teaching point of view, you're trying to get graduate outcomes at the least amount of costs.*

–TE.10

This paradigm should not be confused with pure cost minimisation. In a broader sense, productivity may be framed in terms of wholesale input minimisation, but as discussed in Chapter 3, such a framing is not relevant to higher education, notably public universities in Australia.

*The financial bottom line, often touted by ministries or consultants, is the most pernicious. Financial bottom line denies the role of mission.*

–TE.9

### **7.3.1|d. Process-oriented productivity**

The process-oriented productivity paradigm focuses on key work processes themselves, not the inputs and outputs of those processes. This perspective, which emerged in the interviews, does not have an analogue in the examples from the literature in Table 25. The process-oriented productivity paradigm does not lend itself to technical study or measurement of productivity as it relates to input-output ratios. Process-oriented productivity relies on qualitative judgements of productivity based on inherent qualities and behavioural characteristics of academic work.

*Just looking at inputs and outputs, we're missing a lot of the invisible work about how knowledge work is actually done and undertaken.*

–KS.9

The paradigm also lends itself to a more holistic idea of higher education processes, its mission, and notions of higher education as a public good for society. The process-oriented productivity paradigm takes a long-term view of higher education's most important outcomes, where the fruit of truly productive and high-quality processes would not naturally be visible in short-term outputs.

*The fear is that a productivity indicator in higher education could become the ultimate decision-making factor, as it has become in other industries. How could you ever judge an education until you look back on your career and life and acknowledge the opportunities it has afforded you?*

–TE.3

### **7.3.1|e. Pragmatic productivity**

The pragmatic productivity paradigm could be described as a merging of Ghobadian and Husband's (1990) 'technological concept' and Pritchard's (1995) 'performance approach'. The paradigm acknowledges that process innovations can be made either to reduce slack with the use of inputs or to increase the value of outputs. The paradigm maintains that value should be observable and characterizable, but also that there are deeper, non-financial ways to characterise value in pursuit of mission. This paradigm is concerned with systems innovation and people-focused solutions to achieve demonstrable progress toward strategic objectives.

*Consider university engagement with industry. Productivity quite often is related to economics... But we don't always want to put [engagement] in an economic relationship. There is something else of benefit that would be generated by integrating with organizations in a way that creates value for them and for us.*

–KS.8

*You need to think systematically about issues to even know what is the information we need... and then we need to go out and collect that information. Given the funding constraints, given the pressures where under, what's the evidence base to evaluate strategies?*

–KS.7

### **7.3.2 Model Constructs Assessment**

The construct assessment verifies rigorous definitions of constructs and identifies relationships between the constructs and their real-world instances. Thematic analysis of discussion about model constructs within interview transcripts serves as the foundation of evidence for this assessment. Key findings from the output evaluation are listed in Table 30 and explained in more detail below.

**Table 30: Model construct assessment findings**

| <b>Finding 1</b>                                                                                                                                  | <b>Finding 2</b>                                                                                                     | <b>Finding 3</b>                                                                                 |
|---------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------|
| <b>For representing production complexity</b> , separate productivity estimates in conjunction with joint productivity estimates would be useful. | <b>Direct capital service estimation</b> from empirical data in financial reports was received well by participants. | <b>The value, impact, or quality of research output</b> should be expressed with empirical data. |

For production complexity, the current iteration of the productivity measurement model estimates the joint production of education and research outputs. There are two principal reasons for this: (1) in the first adaptation of the NRC model, APO country participants all expressed interest in measuring research productivity and education productivity, and (2) financial accounting methods in most countries, including in Australia, do not disaggregate expenditure by academic function. Interview participants, however, pointed to troubles with assumptions of joint productivity. Although institutional expenditure on research and teaching is not disclosed consistently in the public domain, both key stakeholders and technical experts spoke of cross-subsidisation practices within institutions of research activities with revenue from education. The implication is that, not only are education and research separable, individual institutions often investigate their research and education productivities separately. Study participant KS.6, a university vice chancellor, mentioned that he had never considered a performance indicator for his own institution to measure education and research performance jointly. “In terms of a composite indicator, I’ve never tried to do that before.”

The model adaptation using Johnes’ (2006) methods to estimate university capital services was received well by participants. In the absence of any other recommended source of data on estimating the value of annual capital flows for universities, the adaptation appears to hold potential for deeper insight on the effectiveness of university resource allocation. One interviewee elaborated.

*It is useful to think about the breakdown of expenditure, particularly capital and recurrent. If you want to break things down empirically it's about labour costs, recurrent equipment costs, and capital. Certainly, capital expenditures have taken off. In Australia it would be useful to just describe these trends.*

–KS.3

Participants finally stressed the importance better capturing some aspect of research quality, value, or impact. They noted that the ‘weighted publications’ metric published by the Department of Education’s Higher Education Research Data Collection (HERDC) initiative was insufficient. The observation further clarifies why HERDC discontinued this indicator as a performance indicator for the system. While participants stressed that using citation data from bibliometric aggregators has many caveats, they also acknowledged the role citation information plays for institutions and for individuals. Among the participants who spoke to the topic of capturing the value or impact of research, all were supportive of using citation data from bibliometric aggregators like Scopus.

*I don't think [using Scopus] is going to cause you any problems from a research design point of view. It's not perfect, but it's used so much.*

–KS.7

*How do you measure research quality? It's always going to be difficult, but look at citations. There are problems, but it doesn't mean you don't look at it.*

–TE.4

### 7.3.3 Construct Associations Assessment

The construct associations assessment involves checking for rigorous definitions of the associations and the nature and appropriateness of the associations. Institution by institution measurement results provide insight on construct associations and how they represented input-output dynamics of Australia's universities. Key findings are provided in Table 31 below.

**Table 31: Construct associations assessment findings**

| <b>Finding 1</b>                                                                                                             | <b>Finding 2</b>                                                                                                                                |
|------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Productivity threshold estimation</b> was validated by participants as a useful tool for interpreting measurement results | <b>Input smoothing</b> over three consecutive years was validated by participants as an acceptable representation of input-output relationships |

The calculation of productivity thresholds was explained in the previous chapter, and Figure 20 above illustrates the range of plausible productivity portrayals a given institution might exhibit depending on how different data elements are emphasised in calculation. Participants overwhelmingly found this to be a helpful tool for interpreting the data and measurements. Two interviewees elaborate.

*In the end, trying to tease out what's causing differences [in productivity]—policy implications—that's important. And maybe your sensitivity analysis tells you a fair bit about that... I'd spend a bit of time on the different weights.*

–KS.3

*I don't like seeing output value weightings stuck at 50-50 because I don't believe that's the case. You need methods of being able to clearly and systematically assign the weights.*

–TE.1

Input smoothing and linking outputs to three consecutive prior years of inputs was also explained in the previous chapter. This was another aspect of the model that was validated by participants. Participant KS.5 centred discussion around short term implications of geo politics that may influence measured productivity but may not be indicative of true changes in an institution's production function. Other participants focused on the inherent lag in in higher education between when decisions are made and when results can be observed.

*The nature of not understanding lags or immediate implications of making changes is a major constraint when you compare higher education to business. Smoothing your inputs is good... Research data is also often lagged.*

–TE.10

### 7.3.4 Model Boundaries Assessment

The assessment of model boundaries explores limitations of the model and its scope for insight. This section focuses on institutional results generated by the model and how it could better capture input-output dynamics within the Australian system. Table 32 summarises findings from the evaluation.

**Table 32: Model boundaries assessment findings**

| Finding 1                                                                                | Finding 2                                                                                       | Finding 3                                                                                                                              |
|------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------|
| <b>The real value of money</b> and currency inflation is not accounted for in the model. | <b>Key aspects of academic work</b> are not accounted for with the current input-output framing | <b>Productivity change estimation</b> in isolation may be insufficient for capturing productivity dynamics across diverse institutions |

A limitation raised during the interviews is that the model does not represent or control for inflationary effects on currency or the ‘real value of money’ over the period examined. Differences in opinion arose among participants about the need to adjust financial figures for the real value of money. The debate is also present in the scholarly literature. Yousif and Dale (1990) explain that decisions to adjust or not to adjust for the real value of money have trade-offs with respect to the accuracy of the productivity portrayal. On one hand, rising prices and currency inflation are observable realities in most economies and should be accounted for and controlled for. Measured changes in institutional productivity should be attributable to changes in an institution’s production technology or key work processes—not input or output price fluctuations. On the other hand, real value for money adjustments are based on estimates, and adjusting figures changes them from their originally observed and reported values. This also runs the risk of creating inaccuracies (Yousif & Dale, 1990). One participant recommended against adjusting financial figures.

*There’s no point [in making the adjustment] because everyone’s working in the same currency and environment... as long as you’re using the same method to measure everyone, then you’re going to have a consistent comparison.*

–TE.1

TE.3 and TE.10 provide the counter perspective. They emphasised the importance of upholding the economic convention of adjusting for the real value of money in the productivity calculation. The concern stems from the fact that the productivity ratio in the current measurement model uses financial inputs but not financial outputs. The implication is that the productivity ratio will be greatly affected by the adjustment because only figures in the denominator of the productivity ratio experience the adjustment. In the for-profit sector, Yousif (1990) demonstrates, unsurprisingly, that productivity ratios are largely unaffected when both the numerator and denominator are adjusted for the real value of money using

similar methods. With the current model in this research, however, not accounting for inflation runs the risk of presenting low productivity change rates that are not representative real monetary inflation in the Australian economy.

As discussed with the process-oriented productivity paradigm above, multiple participants raised the issue that the current model is not able to capture key aspects of academic work. Participant KS.9 emphasised that funding does not translate directly to academic outputs. There are hidden elements, such as network building and the time spent applying for grants. Participant TE.3 mentioned the significant time that academics spend on research training with their post-graduate students that do not always culminate in a measurable output for themselves. Adaptations to the model with respect to these limitations are discussed in Section 7.4.

Finally, a limitation of the model was raised with respect to the reliability of productivity change measurements across a diverse set of institutions. As noted in Chapter 3, when comparing multiple organisations, indexing methods for productivity change are prone to distortions in their productivity portrayals when a baseline for productivity levels is not well-understood at the beginning of the period investigated. Multiple participants raised the issue of diversity among Australian universities and the care that must be taking when comparing their performance portrayals. Participant KS.6 explained the difficulty of establishing a reliable set of peer institutions to which aspects of performance and performance improvement can be compared.

### 7.3.5 Revelation, Novelty, Importance Assessment

The assessment of revelation, novelty and importance focuses on phenomena revealed by the model or uses of the model that were either not previously known or provide deeper insight into phenomena not fully understood. This section summarises many contextual insights provided by participants and illustrates those insights with several indicative quotes. Key findings from the assessment are summarised in Table 33. Explanations of the findings are grouped with respect to (A) novelty and revelation and (B) importance. The discussion about importance highlights responses to the final two questions on the interview form—Appendix E—regarding key areas for improvement to Australian university productivity and constraints to improvement.

**Table 33: Revelation, novelty and importance assessment findings**

| <b>Finding 1</b>                                                                                           | <b>Finding 2</b>                                                                                                     | <b>Finding 3</b>                                                                                                                                                                  |
|------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>The model reveals</b> university productivity trends with a greater level of nuance than prior methods. | <b>The novelty</b> of the model lies in both the indexing approach and the systematic data selection for measurement | <b>The importance of model results</b> lie in its potential to provide inroads for further analysis and decision-making about policy and practice concerns specific to the system |

### **7.3.5|a. Novelty and revelation**

Reactions to the model were overwhelmingly positive—even from participants whose responses largely reflected the ‘process-oriented productivity’ paradigm, which is largely sceptical of input-output measurement.

*We haven’t even had a relevant productivity model in the sector, yet the sector is hundreds of years old. To be able to get something, even if it’s looking at the data measures as proxies for the actual work, at least it’s a start. And then you can become a bit more nuanced down the track. So, I think what you’re doing is a really important thing.*

–KS.9

A former economics faculty dean of a Group of Eight institution remarked,

*I think you’re onto a really good thing. I was thinking, I wish I’d thought of it first. I went to courses on DEA many years ago. It doesn’t seem to give you much. With your model, you can delve more into what’s causing the trends, and that’s what’s important, the ‘why’. What are the characteristics?*

–KS.3

An academic staff member at another institution added,

*We’re not producing cars [in universities], but very few people are. The idea that you don’t look at the productivity [of universities] is unacceptable. It’s a major export. It’s a major part of the economic activity and the identity of what Australia is.*

–TE.4

Another participant who served as chief operating officer at a different institution added,

*I have sympathy for the challenge of findings ways to promote and apply production theory and productivity measurement for use in higher education. That is, to apply a different level of rigor to people’s thinking, which they should be thinking about.*

–KS.5

### **7.3.5|b. Importance of the model**

The importance of the model, as expressed by participants, relates to the informative potential of measurement results to provide insight or inroads to further analysis on key contextual concerns in the system. The importance of the model is also framed in terms of generating key information about universities that could influence funding and resource allocation decisions at the nation policy level. One participant explains that distorted data on university productivity can serve as an unwarranted disincentive for adequately resourcing the system with public funds.

*...that report [on higher education] from [company X] is a really good reason why you should be doing this sort of work. [The report] is purely focused on balance sheet measures and ROI. It says, ‘universities are a wash for cash, so they can afford to be reined in’ without any sense about what the efficiency is,*

*what the outputs are. So, given that report is out there and influencing policy, it's really important to do efficiency studies properly.*

–KS.7

Another participant elaborates that the Australian government is constantly seeking better data and stronger rationales for its higher education funding policies.

*To understand the public benefits [of higher education] is the hard bit... In an Economics 101 kind of world, you look at private wage benefits and economic productivity gains from the marginal person getting a degree... this might inform decisions to subsidise certain types of degrees over other... But if there are other positive externalities associated with higher education, then by definition there is an undersupply of all higher degrees... We need better measures of outcomes.*

–TE.2

The importance of having valid and reliable productivity measures also regards informing stakeholders internal to higher education about the problems that concern them and the areas they want to improve. The final two question of the interview form prompted participants to speak about areas of inefficiency in higher education where measurement could help to benchmark progress toward improvement. A Deputy Vice Chancellor at one institution summarised the issue.

*Waste is defined as activity that doesn't generate direct value for the client that you're there for... So if you take that definition of waste and you apply that to a university, then everything that we are doing here should be directly related to generating a superior student experience or a superior research outcome... and if you look at it from the point of view of the students or the users of research services, then there is an enormous amount of waste in a university.*

–KS.8

Further, four distinct areas of importance arose frequently with respect to what universities could improve with better information. Table 34 shows the results.

**Table 34: Participant-identified areas of inefficiency for universities**

| Improvement Area                          | Participant Frequency | Response |
|-------------------------------------------|-----------------------|----------|
| Utilisation of facilities                 | 9                     |          |
| Student attrition                         | 8                     |          |
| Academic workload models                  | 8                     |          |
| Administrative and non-academic functions | 8                     |          |

Indicative participant quotes help explain these topics. With respect to facilities usage, participant KS.5 noted that most universities in Australia at any given time are using only 30% of their facilities productively. The participant explained that the figure is an estimate based on assessments from the Tertiary Education Facilities Management Association (TEFMA). Regardless of the precision of this estimate, however, the participant emphasised the difficulty in creating a data-driven strategy for facilities usage.

*Timetabling in universities is difficult to carry out efficiently. It's hard to predict the final enrolment in a subject, and it's even more difficult to predict attendance after enrolment. So, you'll sometimes see cases where the head of a subject will book individual rooms for five sections before the total number of students in the subject is known. But then as the semester goes on, only three of those rooms are being utilised. There are now two empty rooms still booked in the system, and another subject with low attendance will end up in a large lecture hall because no small rooms appear available in the system.*

–TE.6

Another participant added,

*I've seen a former VC attempt to introduce a summer term... but the real problem was Christmas. And this doesn't happen in the northern hemisphere. So, you have this big problem that you can't really use December. You could shift the whole year around, but whether you could use facilities for other purposes, you always face the question about how facilities could be better used over the summer.*

–KS.3

With respect to student attrition, one participant summarised the issue.

*...one of the big areas [to improve] is attrition. It's not valuable to anyone involved. It doesn't help the student or the tax payer... and [it's] not a help to the university either. That's probably the number one thing. Students going to university, incurring massive debt, and then not even getting a degree at the end of it.*

–TE.8

Academic workload models represent a complex area to improve for Australian universities. The lack of quality information about the time and effort spent on teaching and research, along with wide variation in workload expectations between universities and within universities impose difficulties for understanding what workload arrangements are best for what contexts. Participants noted that there exists little precise empirical evidence on how staff use their time daily. They further spoke to questions of whether academic staff should be expected to engage in both research and teaching or whether they should specialise. Participants admitted that there are likely no universal answers or solutions to how academic staff members should use their time. Not having quality information on the topic, however, likely means that productivity opportunities are being missed. One participant elaborated.

*For some institutions it's very broad. They'll say that the academics spend 40% of their time on research, 40% on teaching, and 20% with admin and coordination. But it varies completely by institution... We have some clients that have one profile across the university, apparent in their EBA [enterprise bargaining agreement]. Others will have performance agreements that are signed off by every individual academic.*

–TE.6

Administrative processes and the proportion of administrative staff to academic staff is marred with controversy. Participants explained that the prime issue to overcome often

relates to the redundancies that are allowed by a devolved governance structure. One participant from a Go8 institution explained.

*One of the reasons that this university and other universities are [reducing non-academic staff] is because we have a devolved governance structure, where individual faculties are very powerful. Faculties were free to do their own hiring and internal structural decisions. There was no coordinating oversight, so faculties were able to grow inefficiently.*

–KS.7

Another participant from a regional university added,

*When I first started... we had two layers of administration. We had a central administration and faculty administrations. There was duplication and competition between the administrations.*

–KS.6

The issue of non-academic functions was raised most frequently in a negative light, but the increasingly important roles of non-academic staff members were also acknowledged. The participant from above continued.

*The nature of the business has become much more complex over time... You need people who are able to think systematically about issues. It's not just sitting in your office, scratching your head, saying 'wouldn't it be a good idea if we did this?'... These sorts of things can't be done by a dean or a senior professor in the 15 minutes of the day when they're not thinking about research or other things. So, you do need that specialist management, and you need it to be professionally supported.*

–KS.7

### **7.3.6 Evaluation Conclusions**

A prime purpose of the model evaluation has been to establish the evidence needed to frame and justify actionable revisions to the measurement model. In addition to this, Australian higher education key stakeholders and technical experts raise subjective, anecdotal, empirical, and objective trends concerning Australian higher education performance and productivity. They successfully framed numerous factors that are interlinked with university productivity and its measurement. The discussion about areas for improvement revealed several trends that concern institutional leaders, industry professionals, and academic and non-academic staff members alike. Further discussion about big-picture issues for Australia's higher education productivity are discussed in chapter nine. The following paragraphs explain how technical and contextual feedback from participant interviews highlight targeted aspects of the model for revision.

The issues raised during the evaluation reveal multiple anchor points in the measurement model on which to focus for improvement. The challenge is to prioritise which pieces of evidence to translate and incorporate into technical model redesign. Choices naturally rest on the design science model assessment criteria. The assessment criteria provide categories of evidence that support only certain types of model revisions. Revisions, however, must also

reflect the diverse views of the 20 interview participants. Thus, decisions for how to revise the model also aim to equitably represent participant interests. The productivity paradigms identified also serve as an organising principle in how to incorporate evidence for model revision. Adapting the model to accommodate range of productivity conceptualisations is also an objective of model redesign.

The next section details six separate adaptations to the model and supports them with evidence from the model evaluation, as well as with reference scholarly literature. Rationales for model revisions also include additional pieces of evidence from the participant interviews not explicitly listed under the four design science model assessment criteria. A prime criterion for deciding which improvements to make is that criticisms and suggestions from participants operating under each productivity paradigm identified above must be entertained. For the model to hold potential for capturing multiple interests, those interests much be incorporated into model constructs and construct associations

## 7.4 Australian Model Enhancement

### 7.4.1 Model Redesign

This section synthesises all findings from the model evaluation discussed above. The insight generated is directed toward model improvement and redesign. The section defines and justifies all construct and construct association redesign specifications for the final iteration of the model. The purpose of the final iteration is translate the contextual, experiential and technical feedback of Australian stakeholders into a model that is fit-for-purpose within the local higher education system.

Restating details from the introductory section of this chapter, six adaptations to model constructs and construct associations are identified to redesign the second version of the model. The adaptations include:

1. Adjusting estimates for currency inflation and the real value of money
2. Tracking individual data element trends to associate with productivity estimates
3. Incorporating staff labour inputs to track single factor productivity
4. Calculating separate education and research productivity
5. Scaling productivity change trends with absolute productivity approximations
6. Restructuring research outputs for increased internal consistency

The adaptations are discussed in sequence in the sections that follow. Each adaptation is presented in terms of how it addresses issues raised in the evaluation above. Table 35 gives the updated input-output framework for the redesigned model. Figure 24 provides a visualisation and schematic to elucidate model construct associations, with redesigned elements highlighted in red.

**Table 35: Model input-output framework**

| Construct Type | Construct | Variable | Data elements |
|----------------|-----------|----------|---------------|
|----------------|-----------|----------|---------------|

|           |                           |       |                                                    |
|-----------|---------------------------|-------|----------------------------------------------------|
| Input TFP | Labour                    | $L_1$ | Academic staff salary and benefits                 |
|           |                           | $L_2$ | Non-academic staff salary and benefits             |
|           | Capital Services          | $K_1$ | Depreciation and amortisation                      |
|           |                           | $K_2$ | Repairs and maintenance                            |
|           |                           | $K_3$ | Interest payable                                   |
|           | Intermediaries            | $I_1$ | Grants and scholarships                            |
|           |                           | $I_2$ | Administration and other expenses                  |
| Input SFP | Teaching Labour           | $L_3$ | FTEs Teaching Only Staff                           |
|           | Research Labour           | $L_4$ | FTEs Research Only Staff                           |
|           | Teach & Research Labour   | $L_5$ | FTEs Teaching & Research Staff                     |
|           | Non-Academic Labour       | $L_6$ | FTEs Other Function Staff                          |
| Output    | Equivalent FT load Course | $E_1$ | Equivalent FT student load coursework              |
|           | Coursework completions    | $E_2$ | Number of coursework graduates                     |
|           | Equivalent FT load Res    | $R_1$ | Equivalent FT student load research                |
|           | Research completions      | $R_2$ | Number of research graduates                       |
|           | Publications              | $R_3$ | Total scholarly output                             |
|           | Citation Impact           | $R_4$ | Output in the 75 <sup>th</sup> citation percentile |
|           | Research income           | $R_5$ | Amount of income for research                      |

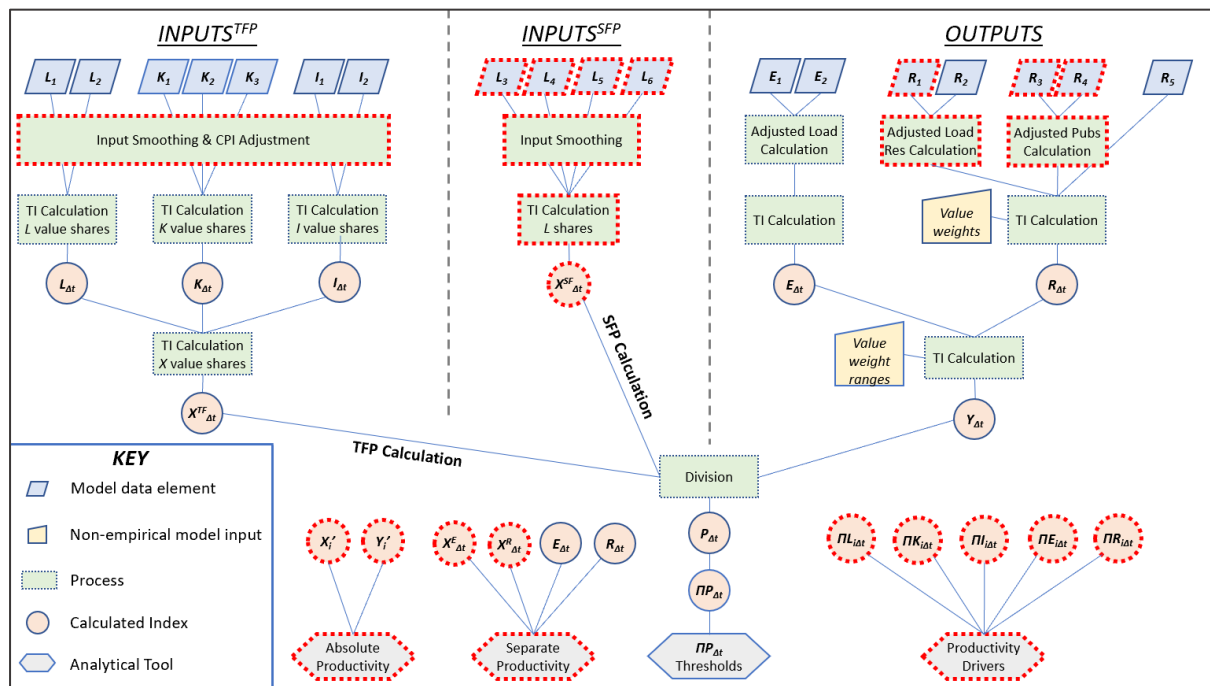


Figure 24: Final model iteration schematic

#### 7.4.2 Real Value for Money Testing

The model redesign adjusts financial variables for the real value of money using the Australian consumer price index (CPI) with a base year of 2007. The adjustment is intended to improve the accuracy of TFP change measurement over the period examined. Based on participant advice and standard practice in the scholarly literature, it is deemed important to at least compare adjusted vs. non-adjusted estimates. Among all potential methods for adjusting for the real value of money across a period—including using a GDP deflator or internal methods—Participant TE.10 recommended using standard CPI data. Yousif (1990) notes that using internal deflators to track price fluctuations for individual inputs and outputs within a given industry is generally understood to produce more accurate results than using

deflators designed to adjust for fluctuations across the whole economy. In the absence of empirical information on higher education input price fluctuations, the CPI can provide a baseline to control from broad inflationary behaviour in the economy. Developing an indicator for Australian higher education input price fluctuations could be a matter for future research. In Figure 21 the current model adaptation can be seen as a Process operating on all TFP inputs.

### 7.4.3 Productivity Drivers Calculation

Assessment findings from section 7.3.5 highlighted the value of the model in its ability to be decomposed into its component parts and illuminate trends in the data elements most associated with final measurement results. The quality of the model was touted as an improvement upon alternative methods for data aggregation. One participant elaborated on the benefit of further exploring data used in calculation.

*One thing that would be useful would be to link back to changes in the nature of the inputs and in terms of what is going on with driving overall trends.*

–KS.3

The phrasing of this quote is what inspired the name of the current redesign option, productivity drivers. The productivity drivers are calculated using the cumulative change operation discussed in Section 6.4.6 of the previous chapter. The cumulative change operation is performed on all data elements used in the model over the period examined to expose relationships between individual data elements and the final productivity index. In Figure 21, the productivity drivers calculation can be seen as an Analytical Tool at the bottom of the schematic.

### 7.4.4 Single Factor Productivity Labour Inputs

Measuring SFP means considering an alternative set of inputs to the financial indicators used for TFP calculation. The SFP indicators include staff numbers and time spent on the broad academic or non-academic functions they are hired to perform. Measuring productivity in this way could begin to address key concerns and assumptions associated with the ‘process-oriented productivity’ paradigm. Section 7.3.1 discusses how the process-oriented productivity paradigm raises concerns that input-output measurement tends to gloss over the nuances of academic work.

*If you look at productivity purely on a financial basis... it leaves out a lot of the invisible work that actually gets done—the academic work—but it’s not counted. Or it’s hard to count. And then it doesn’t get valued.*

–KS.9

Labour productivity was also expressed as crucial in itself.

*Labour is such an important part of the cost structure, and our competitiveness depends so much on our labour productivity.*

–KS.6

With the objective to address process-oriented productivity issues, the final version of the model provides an alternative SFP, labour productivity, estimation to compare against the TFP estimation. The new variables considered are  $L_3$  through  $L_6$  in Table 31, which represent full time equivalent staff figures with respect to the functions they are contracted to perform, ‘teaching only’, ‘research only’, ‘teaching and research’, and ‘other’. The ‘other’ category encompasses staff contracts for administrative and non-academic work. These designations are not precise, but they allow for a better representation of how staff are expected to use their time. The indicators do not solve the issues mentioned in Section 7.3.5 concerning the inconsistencies between institutions for how academic work is tracked and assigned. The FTE input figures do, however, provide a baseline for discussion that speaks to time usage and the effort spent on teaching and research more so than financial inputs.

Measuring labour productivity with staff FTE figures is not unprecedented in the literature, but the current model’s application is novel. Miles et al. (2018) use FTE figures for instructional staff to measure SFP for education in the US, and Abbot and Doucouliagos (2003) used staff FTE figures amidst other inputs to measure multi-factor productivity (MFP) for education and research in Australia. The current application of staff FTE figures allows for labour productivity estimates and also offers an opportunity to examine education and research productivity separately, as explained in the next section. Data on staff FTE by academic function are available at DET (2018d). In Figure 21 the SFP calculation can be seen in the second column of the schematic.

#### 7.4.5 Separate Productivity Calculation

The separate productivity calculation addresses concerns raised in Section 7.3.2 with respect to the limitations of a joint productivity model for education and research output production. While the benefits of a teaching-research nexus may exist, institutionally, education and research are separable functions.

*The teaching and research function. They sometimes have a lot of synergies, but more often than not... they are very divergent. In terms of viewing productivity, there is value in splitting the functions, measuring value on a teaching scale away from value on a research scale.*

–TE.8

Measuring productivity separately is possible because of the SFP calculation described above. The staff FTE figures by academic function afford the opportunity to explore a separate productivity model because data on staff functions allows for linking research staff to research outputs and education staff to education outputs. By contrast, publicly available financial figures used in TFP calculations are disaggregated only by standard accounting categories and not by academic function.

The model specification that allows for calculation of education and research productivity separately is a decomposition of the TI used in the SFP calculation. Details for the decomposition may be seen in **Appendix F**. In the schematic in Figure 21 above, the separate productivity calculation appears as an Analytical Tool at the bottom of the figure.

#### 7.4.6 Education and Research Output Consistency

Three concerns raised during the evaluation are addressed by adapting the model for more consistency with education and research output aggregation. First, in Section 7.3.2 participants noted that current research output data elements lack in capturing the value or impact of research. Second, in Section 7.3.4 participants noted the lack of the model's ability to represent nuanced academic work. Participant TE.3 emphasised unaccounted for time spent on research training that does not always culminate in a measurable output. Finally, one participant noted inherent value of improving the model's internal consistency.

*The more structure you build into a model, the better the results... In higher education there are basic assumptions that you can't always trust because so much of it is variable.*

–TE.9

Specific adaptations to improve internal consistency include two new research output constructs, 'adjusted load research' and 'adjusted publications'. The new constructs improve internal consistency because they each account for quantity and value in the same systematic fashion as is accomplished with the 'adjusted load' construct for education outputs. The following paragraphs detail the adaptations and explain how internal consistency improvements address participant concerns.

First as shown in Table 35 and Figure 24, 'adjusted load research' augments 'research completions'. The new construct incorporates equivalent full-time student load (EFTSL) for graduate researchers. Initial inclusion of research completions in the model was based on its use by the DET as a key performance indicator in funding formulas. 'Adjusted load research' still incorporates research completions and is calculated in the same way as 'adjusted load' for education. In summary, the new construct increases consistency in the model and directly addresses participant concerns about prior versions of the model not accounting for 'hidden work', including research training. For further clarification, The Research Training Program (RTP) is the DET's support scheme for the provision of masters and doctoral programs by research (DET, 2019). Including EFTSL figures for graduate researchers more directly captures the scale of institutional research training efforts than the size of the block grants awarded to the institutions.

The new 'adjusted publications' construct addresses participant concerns related to the prior model's lack of ability to represent the value and impact of research. The adjusted publications construct thus seeks to capture empirical representations of the value-add of scholarly output. First, the adjusted publications construct operates in the same systematic fashion as the adjusted load constructs by adding value indicators to quantity indicators. Adjusted publications, however, is underpinned by a different set of industry-standard empirical evidence. Recall from Chapter 5, work by Sullivan et al (2012), Jaeger and Page (1996) and Park (1999) attributes some inherent value to a student's undertaking coursework in a university on the basis of human capital theory. Using wage information and signalling theory, the additional value-add of earning a degree can be approximated and added to the value of undertaking coursework. The value-add of earning a degree was approximated as equal to the value of one year of coursework.

The adjusted publication construct accounts for the value-add of publishing scholarly output using citation data and field weighted citation impact, as described in Colledge and Verlinde (2014). First, the adjusted publications construct assumes some inherent value associated with the publication of scholarly work, on the basis of Australian university mission objectives to produce knowledge. Using work by Gralka et al. (2019) and Bornmann and Haunschild (2016) as precedent, the additional value add of scholarly publications can be estimated using field-weighted citation impact. The authors acknowledge that citation data cannot capture the full value of research outputs. Further, different fields of study have different citation rate conventions. Using top citation percentiles and normalised citation indicators, however, can provide a valid starting point for estimating the value of research output for efficiency and performance assessment. For the 38 Australian universities in the dataset, a significant number of their publications enter into the global 75<sup>th</sup> percentile of frequently cited papers—that is, many of their papers belong to the 25% most frequently cited in the Scopus SciVal database (Elsevier B.V., 2018). Further, these highly cited articles have a field-weighted citation impact that is approximately double that of their average publications. Hence, the additional value of producing an article that enters into the global 75<sup>th</sup> percentile is accounted for by adding the quantity of those publications to the total amount of scholarly output they produce. The effect is to count the most frequently cited papers twice because of the extra value they generate through their more frequent use and dissemination.

In summary, the adjusted publications calculation is simple, but it rests on consistent principles with those of the ‘adjusted load’ calculation. It is empirically supported, based on scholarly literature and directly addresses concerns of participants. The metric incorporates (A) the total quantity of published scholarly output per institution per year and (B) the amount of those publications that emerge within the global top percentiles of highly cited papers within their academic disciplines. The metric then adds the two quantities together per institution, per year. With respect to Table 35 and Figure 24, the adjusted publications construct is equal to  $R_3 + R_4$ . And hence, an institution that is increasing its highly cited scholarly output will see the increase reflected more positively in its research output index than that of an institution not producing highly cited papers.

#### **7.4.7 Approximating Absolute Productivity with ‘Relative Change’**

Approximating absolute productivity with a relative change estimate addresses concerns raised in Section 7.3.4 about the reliability and interpretability of productivity change results when comparing estimates across the range of Australian institutions. The novel method introduced in this section for estimating absolute productivity also addresses concerns raised in Chapter 3 regarding the opportunities for education efficiency research to better incorporate learnings and practice from broader literature on measurement. The prime limitations of indexing methods are acknowledged in the literature (Singh et al., 2000) and were discussed in Chapter 3, regarding initial productivity conditions per institution. Addressing the limitation also speaks to participant concerns mentioned in Section 7.3.4 about how university comparisons can be misleading when it is known that not all institutions are starting on even ground.

It is important then to distinguish between ‘normative’ productivity change estimated by standard TI calculations and ‘relative’ productivity change that can be estimated if absolute productivity approximations are provided for institutions at the beginning of the period examined. The TI method normalises change by assuming each institution has an initial productivity level of ‘one’ so that its change can be measured reliably with respect to itself. Imagining how much the multiplicative TI estimates could be affected if institutional starting conditions were known with respect to relative absolute productivity estimates. To illustrate why TI’s normalised change trends can be deceiving, consider the hypothetical example given in Box 4.

**Box 4: Normative change vs. relative change**

**Institution A** graduates 20 students on a budget of \$1000. It has a measured absolute productivity value of ‘0.02’.

**Institution A** graduates 10 students on a budget of \$1000. It has a measured absolute productivity value of ‘0.01’.

Each institution experiences positive productivity change by a common factor, doubling their productivity. Institution A will thus have a measured absolute productivity value of ‘0.04’, and Institution B will have a measured absolute productivity value of ‘0.02’.

Even though their multiplicative (or normative) TI indexes would be the same, their additive (or relative) changes in productivity are different. Institution A has increased its productivity by 0.02 units, while Institution B has increased its productivity by only 0.01 units, meaning that Institution A has increased its productivity more quickly than Institution B

TI indexes offer an important solution, however, to the problem of aggregating data elements measured in different units. The TI index can do so because it serves to aggregate only per cent change values for individual data elements, rather than their reported quantities. TI data aggregation occurs only when each element is represented in a common unit of measure, i.e. its per cent change over a designated period. The value weights used in the TI calculation are what systematically control for varying quantities or relative amounts of value per data element. A prime reason with linear programming and DEA methods are used is because it offers a solution to the aggregation of data measured in different units.

Broader data science literature, however, offers several common solutions for rescaling and standardising data to common units of measure, such that the variation in the data is preserved (Said & Torra, 2019). **Appendix G** details how input and output quantities can be aggregated using standard scaling methods. It subsequently shows a method for replacing initial normalising values of ‘1’ with the absolute productivity approximations. When the absolute productivity approximations are use in conjunction with the TI method, then a new, relative productivity change trend can be given, which is more representative of additive or

real productivity change, as discussed in Box 3. The methods in Appendix G also follows principles of internal consistency and use the same output value weighting schemes as the is used for the TI change estimates. That is, it uses the same principles for data aggregation that have been discussed and developed throughout this dissertation. The result produced is thus similar to previous results, in that a range of plausible productivity change portrayals is given based on different valuations of the output elements.

In summary, the difference in the normative productivity change estimates and relative, absolute productivity change methods is that normative change uses TI, relative change estimates use TI in conjunction with initial absolute productivity approximations. Based on the source data, each institution will have a unique level of initial absolute productivity, and further, each institution will have a unique range of plausible portrayals of the initial conditions based on the effects of the range of output value weights tested

**Table 36: Summary of all essential model design and re-design innovations**

| Model Adaptation                          | Description/Rationale                                                                                                                                                                                                                                                   | Research Stage |
|-------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|
| Joint education and research productivity | First TI model in the literature that captures joint education and research productivity. Produces a general institution-level indicator without the need for within-institution accounting practices that separately track expenditures to education and research.     | Stage 2        |
| Non-priced Törnqvist output index         | Novel specifications and guidelines for aggregating non-priced outputs into a single change index when individual data elements are measured in different units. Expands the scope of the model to capture broader institutional phenomena.                             | Stage 2        |
| Adjusted full-time student load           | Adaptation to initial NRC specifications for calculating ‘adjusted credit-hours’. Generalises a key model input to a unit more widely measured across international contexts.                                                                                           | Stage 2        |
| Input smoothing                           | Relates information about inputs from three successive years to outputs from a single year. Key work processes for graduating students and publishing research are usually multi-year efforts. Better captures higher education productivity and authentic performance. | Stage 4        |
| Cumulative productivity change            | Generates an indicator that tracks cumulative change at each time point across a period. Enhances interpretability of results and exposes anomalies in data.                                                                                                            | Stage 4        |
| Productivity threshold estimation         | Generates a range of plausible productivity portrayals based on different emphases on data elements. Captures and parameterises deeper complexity within the data and exposes a broader range of interpretations and interests associated with measured results.        | Stage 4        |
| Separate productivity                     | Innovation to the TI technique for generating                                                                                                                                                                                                                           | Stage 6        |

|                                           |                                                                                                                                                                                                                                                                 |         |
|-------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|
| calculation                               | separate education and research trends. Uses inputs measured in staff FTE units by work function. Allows for more nuanced and dynamic assessment of education and research.                                                                                     |         |
| Adjusted publications                     | Incorporates the principles behind the adjusted load calculation to include citation and research impact information with publication data. Captures notions of research output value and quality and increases model internal consistency.                     | Stage 6 |
| Absolute productivity and relative change | Approximates absolute productivity to allow for relative, as opposed to normalised, productivity change trends. Adds nuance to the overall performance assessment and increases accuracy of change trend gradients when making comparisons across institutions. | Stage 6 |

## 7.6 Summary

### 7.6.1 Focus and Contributions

This chapter has described the rationale and methods for conducting 20 interviews with key higher education stakeholders and experts in Australia. The chapter detailed how interview responses were used as key qualitative evidence in a subsequent evaluation of the measurement model under design in this research. As with the previous chapter, the evaluation was conducted using the four design science model assessment criteria. Interview responses provided insight for adapting the model to its application domain, the Australian higher education system.

Interviews provided a range of deeper insight into performance interests and concerns within the Australian higher education system. While participants helped to guide subsequent technical development of the model, they also provided insight about how measurement results should be interpreted in context. Issues raised by participants form the baseline for how measured productivity results may inform conclusions about institutional performance. Participants, for example, spoke to university productivity in relation to national and institutional policy and funding arrangements, as well as specific operational challenges that face universities across the system. They also demonstrated how different stakeholders within the system conceptualise and understand productivity in different ways. These perspectives helped with the identification of four distinct productivity paradigms associated with the views and interests of higher education stakeholders. On one hand, the diverse views help explain why performance measurement can be controversial. On the other hand, the perspectives help form a framework for understanding contextual strengths and limitations of

quantitative indicators. The mix of qualitative and quantitative findings to this point help in reconciling evidence and interests, and observations and ideas.

### **7.6.2 Considerations and Limitations Arising**

At this point in the research, model adaptations have been made to extensively and to systematically addresses theoretical, conceptual, technical and contextual measurement challenges for portraying higher education institutional productivity. As discussed in Chapter 2, some challenges cannot be completely addressed. First on the list is the ability for the model to capture concurrent quality dynamics of inputs and outputs and their quantities change. Model innovations throughout this research have, however, been designed to address the issue of quality and value in novel ways, and to an extent that surpasses previous attempts in the literature regarding productivity change measurement. While the current advancements are significant, they should not be mistaken as enabling the generation of definitive, stand-alone results. Rather, they shift focus and measurement objectives of a productivity measurement exercise away from that of standard production frontier analyses most common in the literature. The current model allows for increased interpretability of results and stronger insight on a broader range of phenomena, but as with any performance metrics, they should always be cross referenced with parallel analyses of university performance that focus on other dimensions and alternative indicators of performance.

Analysis in the next chapter does not focus on any individual, precise measure. Rather, it focuses on what can be learned from the variability in the data. The analysis includes a suite of findings produced by the measurement model developed in this research. These findings are used as the foundation for a productivity and performance assessment of Australian universities. Each successive productivity portrayal should be viewed as revealing a new perspective on the phenomenon and as adding substance to the overall story of performance. Competing portrayals of productivity may serve to reemphasise the issues raised in previous chapters concerning the reliability of institution-level indicators. Alternatively, the range of portrayals provided allows for a more comprehensive picture of performance.

PART III:

DATA ANALYSIS, SYNTHESIS & FINDINGS

## CHAPTER 8: MEASUREMENT RESULTS AND PRODUCTIVITY ASSESSMENT

*You get into all sorts of nuances when you are measuring teaching and research productivity.*

–Study Participant KS.6  
University Vice Chancellor

### **Abstract:**

This chapter applies the final productivity measurement model to Australian higher education data. The prime contribution of this chapter is to generate evidence for addressing research question Q3. The chapter uses multiple analytical techniques to portray Australian university productivity. Techniques include those for estimating total factor and single factor productivity, estimating research and education productivity jointly and separately, and estimating productivity change and absolute productivity. Measurement results highlight functional dynamics of education and research. Two key findings include: (1) increases in research productivity are driving positive productivity change trends across the sector; yet (2) emphasising education productivity in calculations positions most institutions nearer an efficiency frontier. Findings underscore deep complexities within Australian higher education. The results, however, help characterise the complexity and serve as foundational evidence for subsequent policy discussion.

### **Key Words:**

Education, research, productivity, convergence, divergence

## 8.1 Introduction

This chapter presents findings from Stage 7 of the design science research. The contribution of this chapter is to provide the preponderance of quantitative evidence for addressing research question Q3 on the characteristics of Australian university productivity. Because the final version of the model allows for multiple portrayals of institutional productivity, it is helpful to first summarise how the contents are presented. First, TFP change estimates are compared to SFP labour productivity change estimates. Next, results for the decomposition of the SFP index into separate education and research productivity indicators are explored. Discussion then proceeds to the results of the relative productivity change calculation based on approximations of institutional absolute productivity. A final discussion of results is then presented with respect to deeper trends within the data, as well as emergent patterns across institutions.

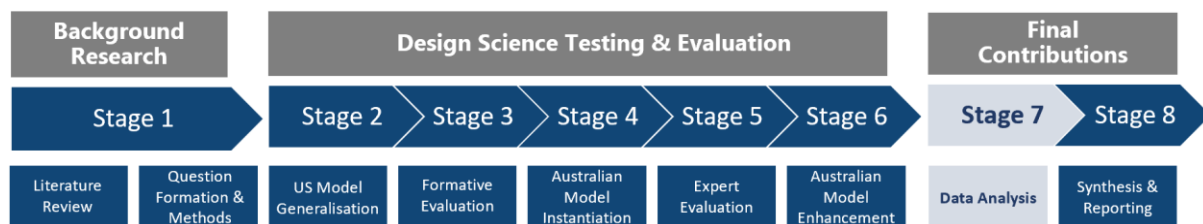


**Figure 25: Research Stage 7**

Important to note are two key changes to the dataset. First a longer period is examined, 2007-2016. To provide a more complete and representative portrayal of system trends, the most recent decade of available data is analysed. Second, as discussed later in Section 8.5, all estimates reflect an adjustment for the real value of money, with all financial variables being measured in 2007 dollars according to the Australian CPI. After testing results with and without inflationary adjustments, it was determined that adjusting for the real value of money provided more accurate results.

It is thus helpful to recognise that the value of empirical findings in this chapter is neither increased nor diminished with tests of statistical significance. While results address research question Q3, the question was formed as intentionally broad. There is no strict hypothesis testing in this research. Further, in the case of exploratory analysis a strength of the TI method is that no statistical inference is involved in the calculation. Productivity estimates directly reflect source data according to model specifications. There is no need for calculating confidence intervals. Rather, sensitivities to measurement results are reported directly by showing alternate results based on different value weightings of data elements. Further, it is both assumed and verified based on the literature review, model assessments, and stakeholder interviews, that variations in institutional sensitivity to alternative estimation techniques do not occur by random chance, but rather occur because of real and observable contextual and

operational differences between institutions. To further elucidate why calculating statistical significance of results is irrelevant for the current analysis, consider the following.

1. Results do not attempt to reject a null-hypothesis associated with interaction effects between specific variables. Rather the analysis is exploratory and descriptive.
2. Results are not derived from a random sample to make inferences about a broader population. Rather, the entire population of Australian universities is measured.
3. Analysis does not involve curve-fitting or statistical regression. Rather, all estimates represent the full variation of source data based on specified value weightings of different data elements.
4. Analysis does not test treatment effects on pre-specified groups. Rather, emergent patterns and classifications are reported.

As explained and demonstrated throughout this dissertation, the state of productivity measurement and assessment is not at the point where measured results are enhanced or benefit from claims of statistical significance. On the contrary, claims of statistical significance serve only to give undue credence to results whose uncertainty is affected primarily by unmeasured, non-statistical matters. Data inclusions and exclusions, performance contexts, politics and stakeholder interests give rise to the most relevant and important uncertainties in measured results. This is why the presentation of results below covers a variety of different interpretations and portrayals of the data. This practice serves to address such non-statistical uncertainties inherent in university performance metrics—even though those uncertainties are difficult to quantify. The transparency of all data treatment, selection and calculation delineated throughout this dissertation should instil sufficient confidence in the fidelity of the numeric indicators presented below.

For expedient interpretation of the results that follow, it is helpful to recall the meaning of key variables in the model. For context, it is also helpful to observe the individual changes in the variables across the system over the period examined. Table 37 presents the TI value for each variable aggregated across all 38 universities in the dataset. Results in the sections below incorporate these indexes, weight them and aggregated or disaggregate according to the specifications described in the previous chapter. Most notable are increases in teaching only FTE staff, total scholarly output, and scholarly output in the top 25<sup>th</sup> citation percentile.

**Table 37: Key variables and system index values 2007-2016**

| <b>Index Variable</b> | <b>Variable Definition</b>             | <b>2007-2016 Index Value</b> | <b>2007-2016 Per cent Increase</b> |
|-----------------------|----------------------------------------|------------------------------|------------------------------------|
| $L_{1\Delta t}$       | Academic staff salary and benefits     | 1.414                        | 41%                                |
| $L_{2\Delta t}$       | Non-academic staff salary and benefits | 1.475                        | 48%                                |
| $K_{\Delta t}$        | Capital services expenditure           | 1.524                        | 52%                                |
| $I_{\Delta t}$        | Intermediaries expenditure             | 1.493                        | 49%                                |
| $L_{3\Delta t}$       | FTEs Teaching Only Staff               | 1.767                        | 77%                                |
| $L_{4\Delta t}$       | FTEs Research Only Staff               | 1.315                        | 32%                                |
| $L_{5\Delta t}$       | FTEs Teaching & Research Staff         | 1.075                        | 8%                                 |
| $L_{6\Delta t}$       | FTEs Other Function Staff              | 1.325                        | 33%                                |
| $E_{1\Delta t}$       | Equivalent FT student load coursework  | 1.369                        | 37%                                |
| $E_{2\Delta t}$       | Number of coursework graduates         | 1.284                        | 28%                                |
| $R_{1\Delta t}$       | Equivalent FT student load research    | 1.308                        | 31%                                |
| $R_{2\Delta t}$       | Number of research graduates           | 1.427                        | 42%                                |

|           |                                                        |       |      |
|-----------|--------------------------------------------------------|-------|------|
| $R_{3At}$ | Total scholarly output                                 | 2.131 | 113% |
| $R_{4At}$ | Output in the top 25 <sup>th</sup> citation percentile | 2.929 | 193% |
| $R_{5At}$ | Amount of income for research                          | 1.223 | 22%  |

## 8.2 TFP Change vs. SFP Change

### 8.2.1 TFP and Change Thresholds

TFP estimates explore education and research output relationships with respect to financial inputs in terms of labour, capital and intermediaries. For the Australian higher education system, and for most individual institutions, the TFP change analysis shows that positive productivity change portrayals across the period depend upon research-output-emphasised calculations. Recall that, as discussed in Chapter 7, examining a range of values for the output weights (a) exposes a range of plausible productivity interpretations for the data under consideration and (b) helps to expose dynamics and trends behind the final estimates. Figure 26 provides an illustration of the TFP change thresholds for the system based on relative education and research valuations from  $v_E/v_R = 0.7/0.3$  to  $v_E/v_R = 0.3/0.7$ .

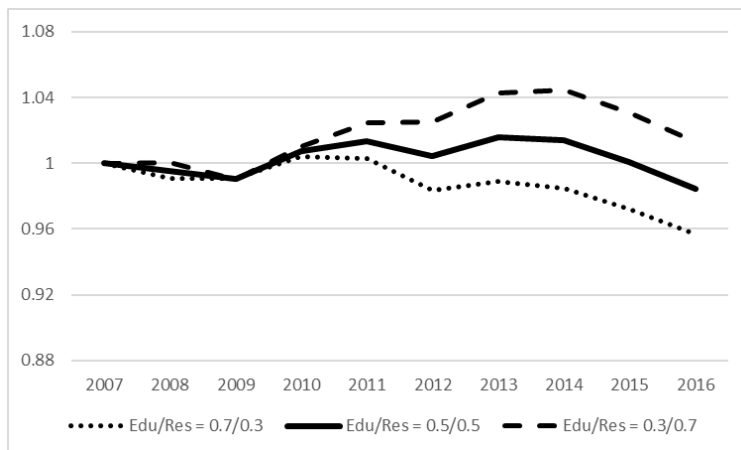


**Figure 26: System TFP thresholds 2007-2016**

The figure shows that with greater emphasis on research outputs, productivity over the period is shown to have grown by just over one percent. With greater emphasis on education outputs, productivity is shown to have diminished to a level of approximately 96 per cent of its value at the beginning of the period. Two clarifications made to avoid confusion or misinterpretation of these results. First the trend shows a change in productivity, not a change in production. It is tempting to view the trend—especially from 2014 to 2016—and contest the results, citing information about demonstrable student and research growth over those years. The trend, however, represents changes in the output to input ratio, not the level of output. Even as system outputs increase, the what the figures shows that the unit output per unit input decreases during the final three years of the period. Second, respective education and research emphasised trends comprising TFP thresholds do not imply separate education and research productivity. They illustrate how portrayals of joint productivity of education

and research vary depending on greater or lesser emphasis on the outputs of each function. Key interests lies in how the threshold of plausible productivity interpretations expands after 2010. The expansion of the threshold signifies that growth rates of research outputs surpasses growth rates of education outputs. Contextually, it would be difficult to disassociate this trend with the impact of global rankings. The year 2010 is a reasonable point in time for strategic policies and initiatives to have begun producing results, in response to the explosion in popularity of rankings systems during the early and mid-2000s, which largely emphasised research performance.

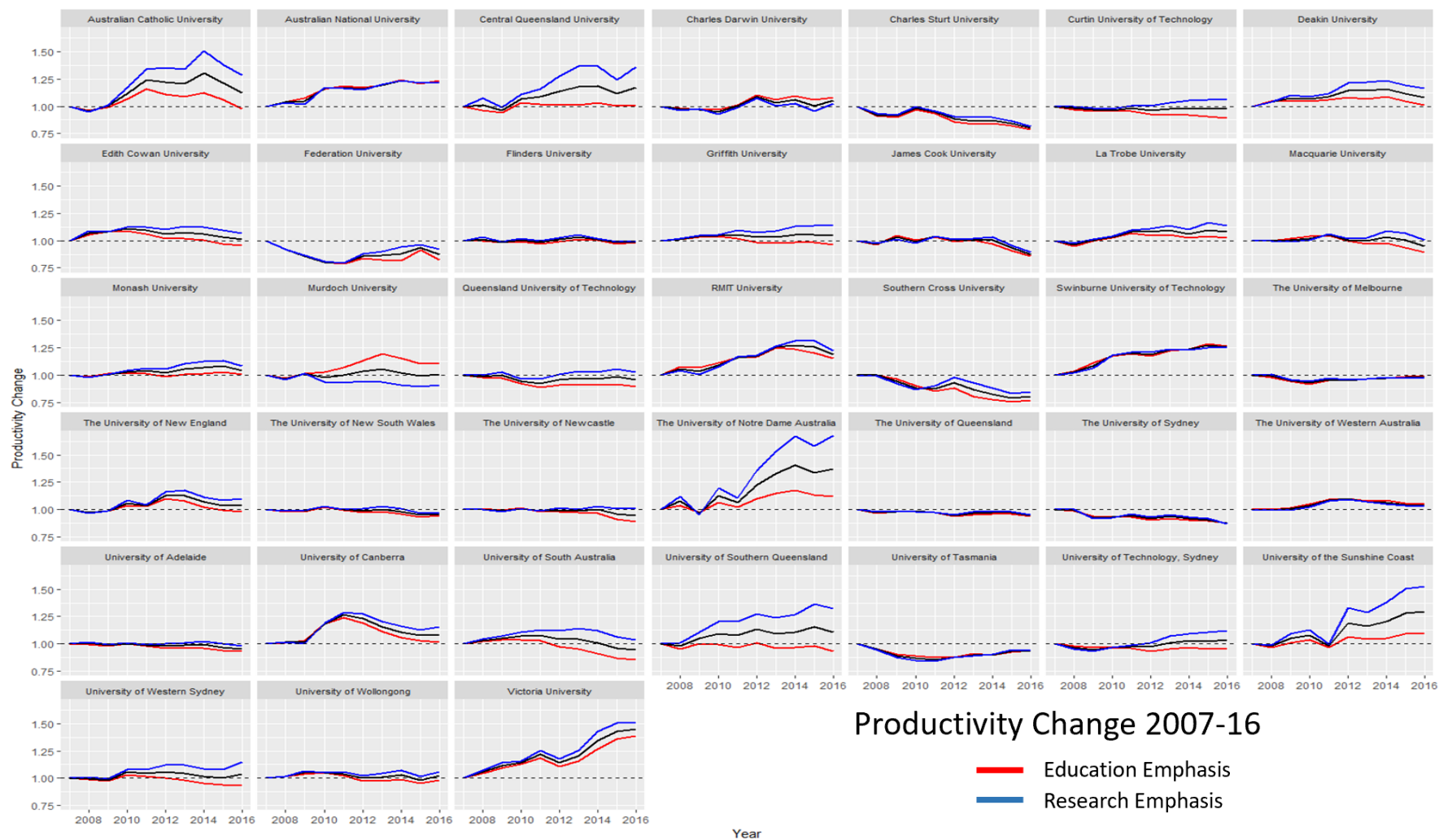


Figure 27: Institutional TFP change thresholds

The system trends from above reflect the trends of individual institutions. **Error! Reference source not found.** explores the dynamics and productivity threshold across all 38 institutions in the dataset over the same period. As above, the figure illustrates productivity change thresholds for relative education and research value weights from  $v_E/v_R = 0.7/0.3$  to  $v_E/v_R = 0.3/0.7$ . The same caveats as above apply, and it is worth reiterating that the trends represent normative change where each institution's initial value is '1' and their growth rate are relative only to their own respective levels of productivity at the beginning of the period. Hence, as explained in Section 7.4.7, the rates of change represent multiplicative change, rather than additive change.

Under most circumstances across the system, more positive institutional TFP change indexes depend on research-output-emphasised calculations. Further among the majority of institutions with an expanding threshold of plausible productivity change interpretations, the expansion appears to begin from 2009 to 2011. As addressed above, the diverging change thresholds are driven by relatively faster increases in research output growth than education output growth. The prime unanswered question about the trends at this point is whether they are being driven by pure increases in research output production, or whether research productivity is increasing. From these measures it is impossible to know. As discussed in Chapter 2, and as discussed further in the next chapter, there is notable cross-subsidisation of research activity with revenues generated from education activities within Australian universities. One can imagine a situation where cross-subsidisation of research inputs might result in a proportional increase in research output. The result, in fact would not be an increase in research productivity, but only an increase in research production with proportionally greater input. Without data on expenditures on academic functions in Australian universities, this is difficult to determine, but the issue may be addressed in another way in the next section with respect to SFP measures of productivity.

First, from a diagnostic perspective, there is more to be learned from exploring the effects of different value weighting schemes on the output data for the 38 institutions within the system. Table 38 shows the effects on relative rankings between institutions by their TFP change rates when different weighting schemes are applied to across institutions. The value weighting schemes are more extreme in the below table than in the threshold analysis above. The decision to include more extreme value weights is because if the objective were to optimise a proactivity change portrayal, as with linear programming techniques, extreme value weighting schemes are usually allowed. The sensitivity of the field of institutions to the treatment of data, however, is precisely the type of insight that is not reported in most linear programming productivity analyses. First take note of the rankings of the top six institutions listed in the Table 38. These six institutions may be positioned in the top 10 or the bottom 10 in terms of their productivity change when different weights are applied to education and research outputs. Further, whether their rankings increase or decrease with education or research emphasis is inconsistent. The variation in rankings speaks to the diversity of operational contexts, strategies and performance priorities. The sensitivity of the rankings illustrates the range of conclusions that could be drawn about institutions when performing a system-level comparative analysis. From another angle, it represents the wide margin for error. Section 8.5.2 explores dynamics among individual data elements to help further explain sensitivities to productivity data aggregation methods.

**Table 38: Institutional rankings' sensitivity to different value weight schemes**

| Institution                            | edu / res<br>0.9 / 0.1<br>Rank | edu / res<br>0.7 / 0.3<br>Rank | edu / res<br>0.5 / 0.5<br>Rank | edu / res<br>0.3 / 0.7<br>Rank | edu / res<br>0.1 / 0.9<br>Rank | Potential<br>climb in<br>Rank |
|----------------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|-------------------------------|
| Murdoch University                     | 4                              | 6                              | 22                             | 34                             | 38                             | ^34                           |
| University of Southern Queensland      | 35                             | 28                             | 9                              | 5                              | 3                              | ^32                           |
| The University of Melbourne            | 8                              | 15                             | 24                             | 28                             | 31                             | ^23                           |
| Central Queensland University          | 26                             | 13                             | 7                              | 4                              | 5                              | ^22                           |
| Charles Darwin University              | 6                              | 8                              | 13                             | 24                             | 28                             | ^22                           |
| University of Western Sydney           | 29                             | 27                             | 18                             | 11                             | 7                              | ^22                           |
| Australian Catholic University         | 27                             | 19                             | 8                              | 6                              | 6                              | ^21                           |
| The University of Notre Dame Australia | 20                             | 5                              | 2                              | 1                              | 1                              | ^19                           |
| Flinders University                    | 9                              | 18                             | 23                             | 27                             | 26                             | ^18                           |
| The University of Western Australia    | 7                              | 9                              | 17                             | 23                             | 25                             | ^18                           |
| University of Tasmania                 | 15                             | 24                             | 33                             | 32                             | 33                             | ^18                           |
| Curtin University of Technology        | 33                             | 31                             | 25                             | 19                             | 17                             | ^16                           |
| University of South Australia          | 34                             | 34                             | 32                             | 21                             | 18                             | ^16                           |
| The University of Queensland           | 17                             | 25                             | 31                             | 31                             | 32                             | ^15                           |
| Griffith University                    | 24                             | 20                             | 14                             | 13                             | 10                             | ^14                           |
| University of the Sunshine Coast       | 16                             | 7                              | 3                              | 2                              | 2                              | ^14                           |
| Australian National University         | 3                              | 3                              | 5                              | 9                              | 14                             | ^11                           |
| The University of New South Wales      | 19                             | 23                             | 27                             | 30                             | 29                             | ^11                           |
| The University of Sydney               | 25                             | 33                             | 35                             | 36                             | 36                             | ^11                           |
| Monash University                      | 10                             | 14                             | 15                             | 17                             | 20                             | ^10                           |
| Queensland University of Technology    | 31                             | 29                             | 26                             | 22                             | 21                             | ^10                           |
| University of Technology, Sydney       | 23                             | 21                             | 19                             | 15                             | 13                             | ^10                           |
| Swinburne University of Technology     | 2                              | 2                              | 4                              | 7                              | 11                             | ^9                            |
| The University of Newcastle            | 32                             | 32                             | 30                             | 25                             | 23                             | ^9                            |
| University of Wollongong               | 13                             | 16                             | 20                             | 20                             | 22                             | ^9                            |
| University of Adelaide                 | 21                             | 26                             | 28                             | 29                             | 27                             | ^8                            |
| Federation University                  | 36                             | 36                             | 36                             | 33                             | 30                             | ^6                            |
| Macquarie University                   | 28                             | 30                             | 29                             | 26                             | 24                             | ^6                            |
| Deakin University                      | 14                             | 12                             | 10                             | 10                             | 9                              | ^5                            |
| James Cook University                  | 30                             | 35                             | 34                             | 35                             | 34                             | ^5                            |
| La Trobe University                    | 11                             | 10                             | 12                             | 14                             | 15                             | ^5                            |

|                               |    |    |    |    |    |    |
|-------------------------------|----|----|----|----|----|----|
| Edith Cowan University        | 22 | 22 | 21 | 18 | 19 | ^4 |
| RMIT University               | 5  | 4  | 6  | 8  | 8  | ^4 |
| Southern Cross University     | 38 | 38 | 37 | 37 | 35 | ^3 |
| Victoria University           | 1  | 1  | 1  | 3  | 4  | ^3 |
| The University of New England | 18 | 17 | 16 | 16 | 16 | ^2 |
| Charles Sturt University      | 37 | 37 | 38 | 38 | 37 | ^1 |
| University of Canberra        | 12 | 11 | 11 | 12 | 12 | ^1 |

Lastly, it should be noted that sensitivities to different value-weighting schemes are multi-dimensional. That is, there exists institution-level sensitivity and system-level sensitivity. Take the University of Melbourne as a prime example. **Error! Reference source not found.** shows that the University of Melbourne's TFP change estimates are insensitive to different output value weighting schemes. Table 38, however, illustrates that its position within system rankings can experience dramatic change because of how other institutions respond to different value weights. The observation cements the fact that single perspectives on productivity, such as a TFP framing, cannot produce definitive results on input-output performance, much less a single treatment of data under a single framing. Therein lies the value of comparing multiple portrayals to triangulate results in hopes of arriving at more robust conclusions. The next section provides SFP estimates.

### 8.2.2 SFP Change and Thresholds

This section examines SFP change indexes. The analysis considers the same set of outputs as the TFP analysis but relates them to inputs measured in quantities of fulltime equivalent members of staff. The staff FTE data considered is segmented with respect to responsibilities in terms of teaching, research and non-academic functions. A prime intention of the analysis is to give an alternative portrayal of productivity that links outputs to non-financial inputs more directly responsible for their production. Figure 28 shows system-level SFP change estimates using the same method of calculating thresholds as above.

SFP estimates give more positive results, and the difference in magnitude between SFP and TFP at the system level is notable. Consider the value weighting scheme with equal emphasis on research and education. The equal weights result in an estimate of SFP growth at just over 11 percent for the period, as compared to equal weights for TFP change showing a two per cent decline. Further, year-on-year SFP growth is consistent across the period. Figure 29 explores institutional SFP change thresholds. Most institutions reflect the system-wide trend exhibiting more positive changes in year-on-year SFP estimates than their respective TFP estimates. Because the output data is the same as the TFP estimates, the expanding thresholds with a more positive research-emphasised estimates are also present. For the same reason, the sensitivity of SFP institutional rankings is largely the same as those of the TFP analysis above. A separate rankings analysis for the SFP measure is not presented because no new insight is revealed.

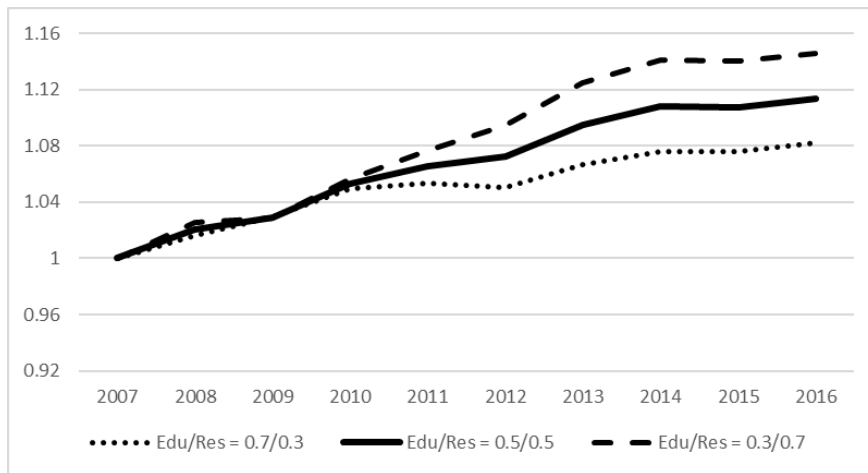


**Figure 28: System SFP change thresholds 2007-2016**

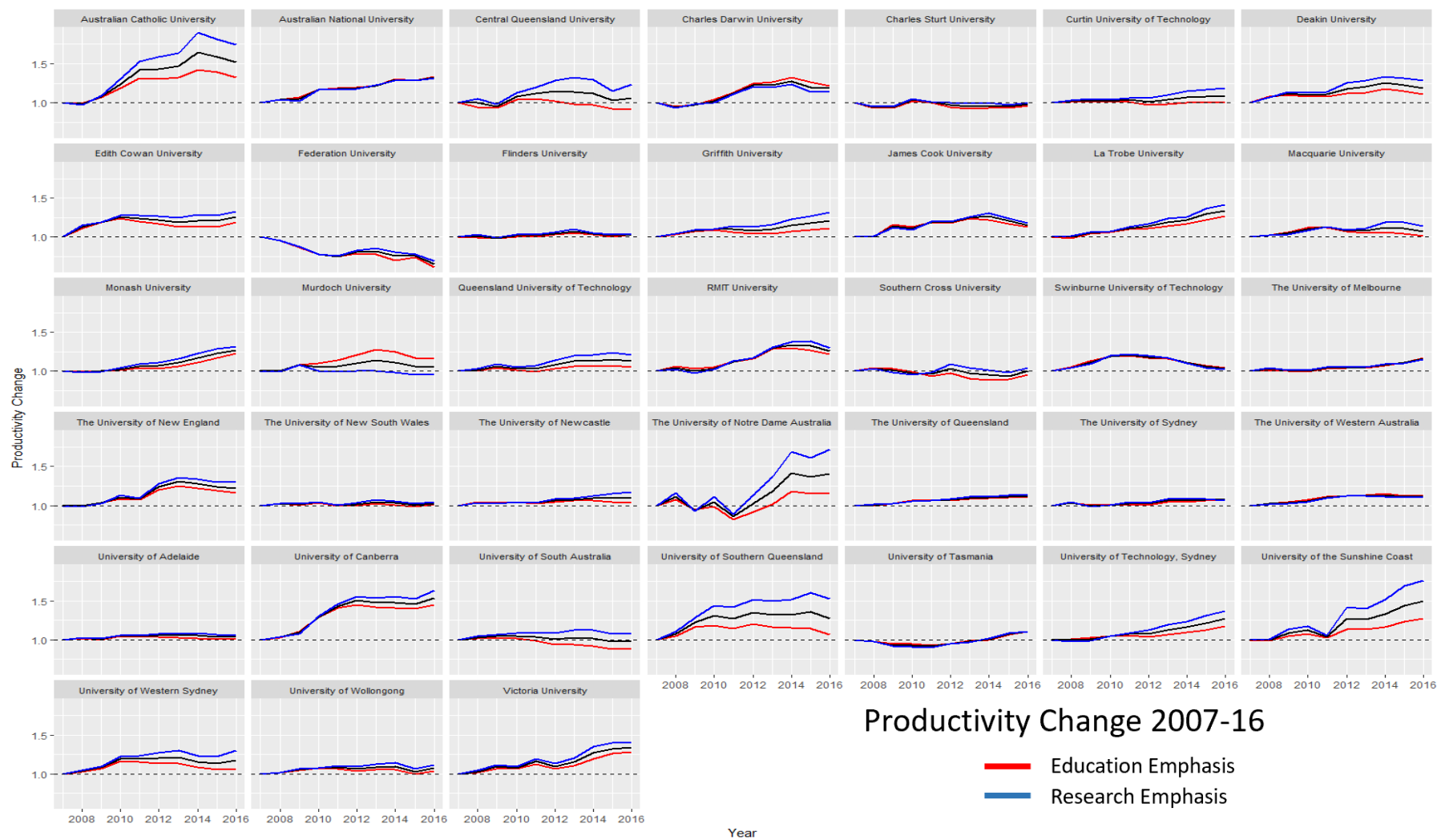


Figure 29: Joint SFP change threshold estimates

The discrepancy between SFP and TFP estimates in this research may have a number of implications, and the finding raises many questions. Several interpretations might be valid. First, with respect to only the SFP results, a quick interpretation might conclude that the gains reflect academic and non-academic staff members using their time more efficiently and effectively. Staff members working the same hours might be producing more output per person per year. Such gains could be attributed to changing staff compositions within institutions. Increasing teaching and research specialisation could also explain the trends. Larger proportions of teaching only and research only staff may sharpen focus on academic objectives. Trends might also be attributed to decreasing proportions of professional staff, where larger proportions of staff directly contribute to academic output. If proportions of professional staff are not decreasing at a sufficient rate to explain the productivity gains, however, it could also be due to increasing trends of non-academic staff performing traditional academic functions.

With respect to TFP trends, the SFP findings could reflect more about non-labour related inputs than the labour inputs. It could be the case that non-labour, capital and technological investments are the prime enablers for staff to execute tasks more effectively and efficiently. Alternatively, if strategic staffing decisions are found to be the prime productivity drivers, results could imply that the non-labour operational expenditures and capital investments are prime slack variables in the system. Answers are not clear, but further implications are discussed in the next chapter, and the findings may open doors for further research.

### **8.3 Separate SFP Change**

The previous chapter discussed how joint SFP change indexes can be decomposed into individual components that give estimates for separate education and research productivity. The decomposition is possible by associating teaching staff with education outputs and research staff with research outputs. The operative assumption for the decomposition is that staff with both teaching and research responsibility spend 50 per cent of their time on each, and the indirect work of non-academic staff is also assumed to equally contribute to education and research outputs. Under these assumptions a system-level analysis of separate research and education SFP change is given in Figure 30.

To frame the significance of the separate productivity estimates, consider the unique shapes of the two curves in the figure above. The trends produced under different output value weighting schemes for joint SFP and joint TFP served only to scale the same underlying information to different extents based on different emphases. A quick glance at the separate education and research productivity portrayals, however, reveals that unique input-output dynamics are being captured in each trend. The two curves represent two independent production functions. As with the joint SFP estimates, both curves show generally positive trends. Research productivity change, however, is more positive and more consistent.

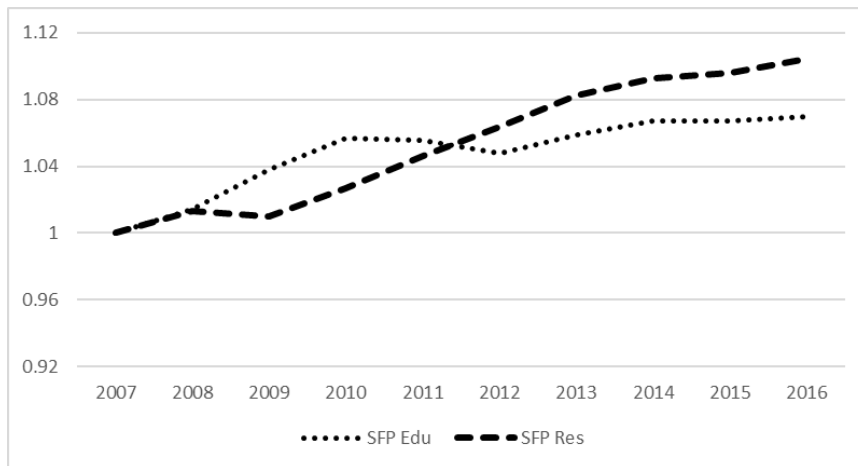


**Figure 30: Separate SFP change system trends 2007-2016**

Figure 31 shows separate SFP change over the period for each institution. The independent nature of the separate research and education trends for each institution is apparent. Results provide mounting evidence, though, for a cohesive story about research driving the sector's productivity growth. Also similar to the analyses above, the beginning of the divergence of the trends generally occurs between 2009 and 2011. That separate education and research productivity trends and education and research-emphasised joint productivity trends tell parallel stories is not trivial. As discussed above, if FTE research staff were growing at a rate proportional to that of research output, then research productivity would not appear to increase. It does appear however, that across the system, institutions are generating more research output per unit input than at the beginning of the period.

To frame the significance of the separate productivity estimates, consider the unique shapes of the two curves in the figure above. The trends produced under different output value weighting schemes for joint SFP and joint TFP served only to scale the same underlying information to different extents based on different emphases. A quick glance at the separate education and research productivity portrayals, however, reveals that unique input-output dynamics are being captured in each trend. The two curves represent to independent production functions. As with the joint SFP estimates, both curves show generally positive trends. Research productivity change, however, is more positive and more consistent.

The separate productivity findings may have several implications for understanding interactions between the two prime academic functions. There appear to be nuanced trade-offs associated with increasing the productivity of either function. While increasing research and education outputs is certainly not a zero-sum game between, there does seem to be a trade-off with respect to improving the education and research production functions. Australian universities appear to be dynamic enough organisations that direct growth or sacrifices in one area do not necessarily lead to proportional drops or gains in other areas. Trade-offs do, however, seem to be associated with the effort and investment required to improve upon production functions. The demonstrated gains in research productivity seem to be at the expense of gains in education productivity.

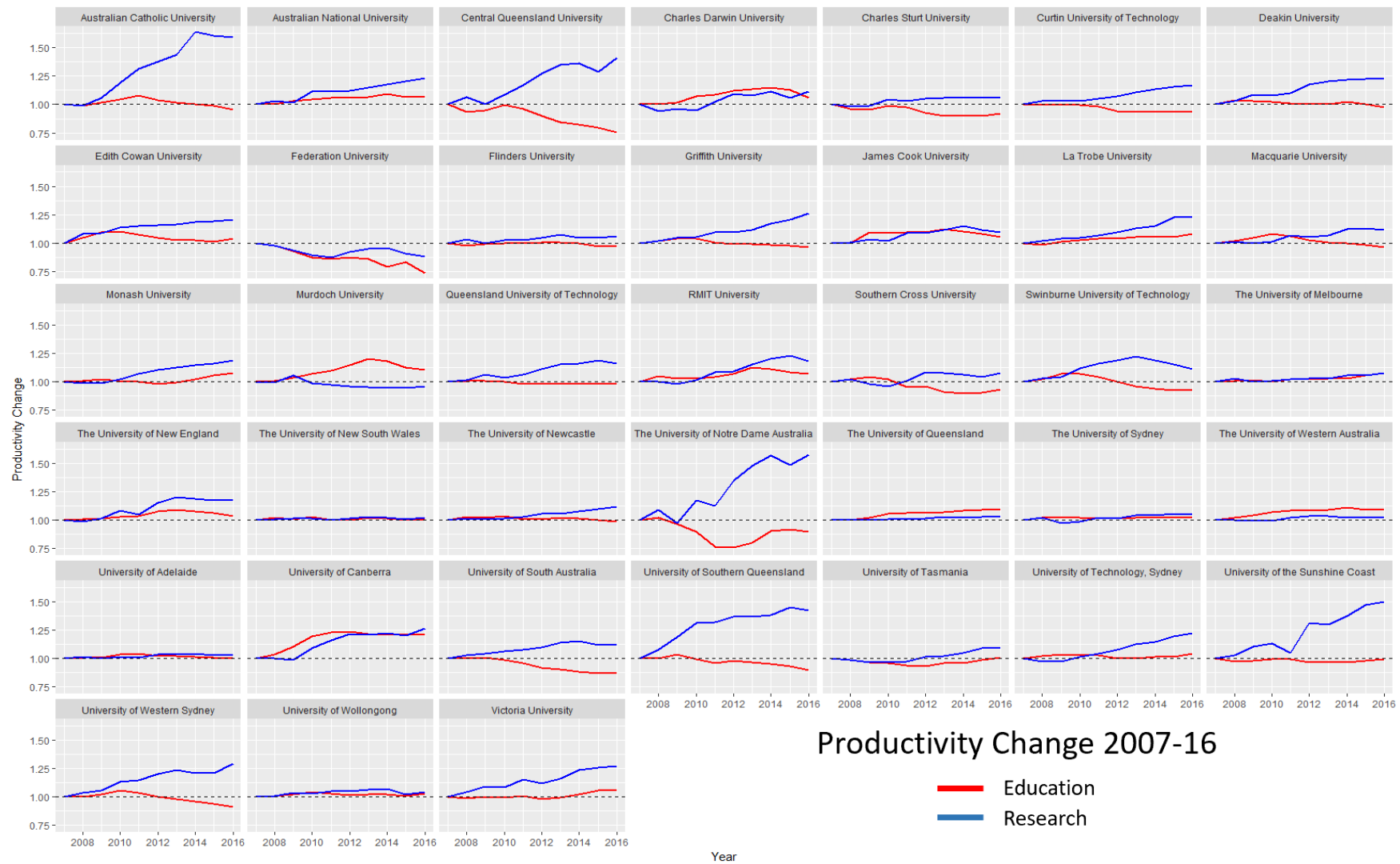


Figure 31: Separate productivity change estimates per institution

There are of course caveats to consider when interpreting this data. Trends could be affected by alternative assumptions concerning the proportion of time and effort that traditional academic staff and non-academic staff spend directly or indirectly contributing to education or research work. Further, there is no measure of the quality of the FTEs. The instructive value of the current analysis, however, is to highlight the extent of the separate behaviour of the distinct academic functions and to shed light on a different interpretation of the data that would not be possible under assumptions of joint productivity.

Over the ten-year period examined, findings suggest that both research output and research productivity have been increasing more quickly than those of education. The story is one of increasing divergence between the two academic functions. In terms of relative patterns of change among institutional inputs and outputs, the story of divergence is one with a strong foundation of evidence. Each institution's productivity change trends, however, are relative only to that institution's productivity at the beginning of the period. The initial relative starting productivities of each institution in the system remain undefined. The next section, however, provides a new perspective on the data using the absolute productivity estimating technique developed during the previous stage of research and explained in Section 7.4.7.

## **8.4 Absolute TFP Change**

This section gives the results of analysing productivity change over the period relative to approximations of absolute productivity. As specified in Appendix G, the TFP change data presented in Figure 26 above is now scaled with respect to initial absolute productivity estimates for the beginning of the period for each institution. All the same data aggregation and analytical principles established in the research are preserved with this calculation, including providing multiple plausible portrayals of the data based on output value weight ranges.

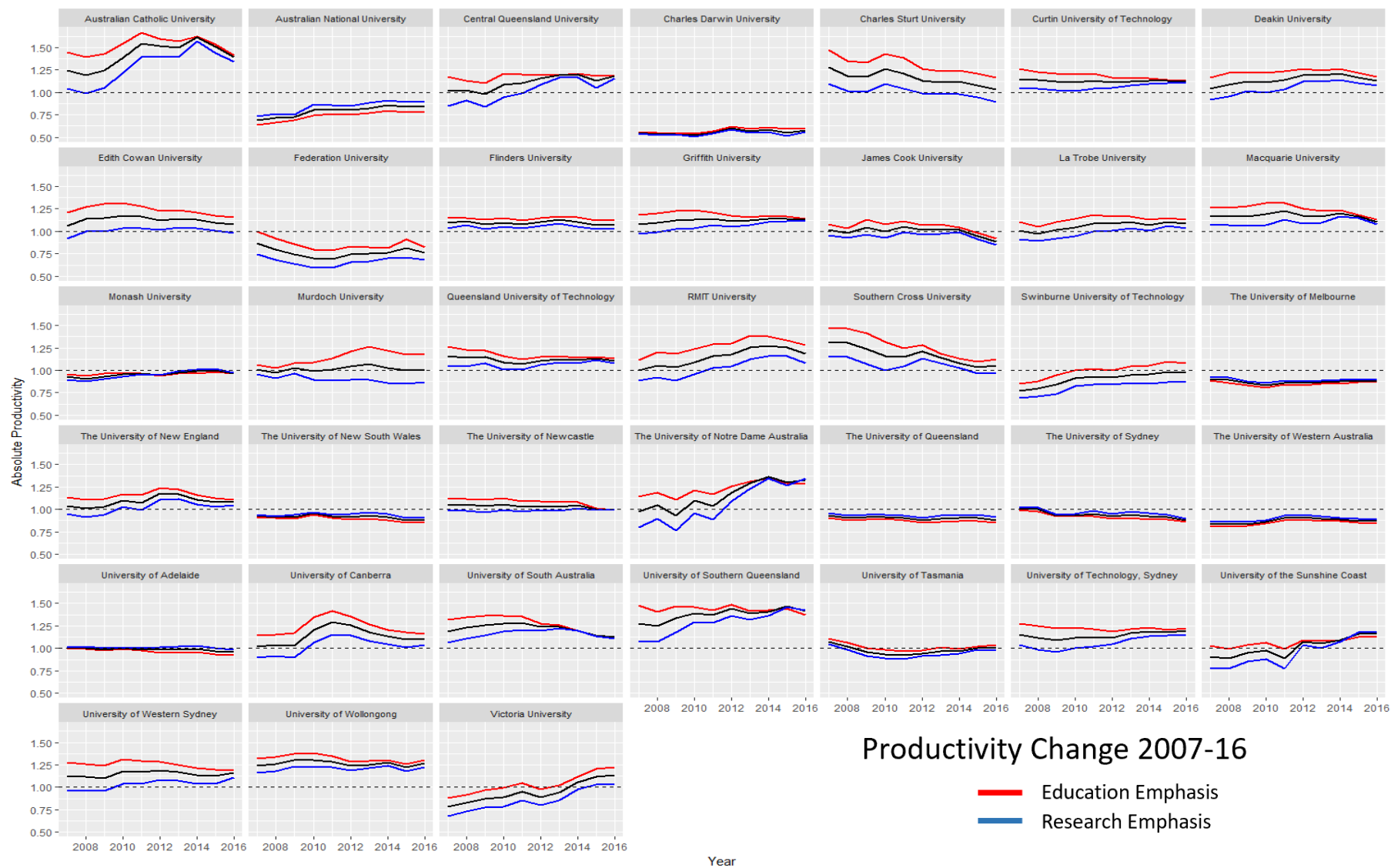
The prime difference is the scale on which productivity change is measured, which is no longer direct TI values. Instead, the new scale is an index for absolute productivity. For consistency in interpretation of results, the scale is also best understood with reference to an absolute productivity score of '1'. The value of '1' on the current scale, however, is defined differently. It signifies a reference level of absolute productivity achieved from dividing the maximum observed output for a single institution by the maximum observed input from a single institution. The input and the output values that compose the index do not necessarily come from the same institution. The reference point is intended to signify something of an 'average' level of productivity for the largest institutions in the dataset. The value of '1' itself is trivial and is assumed to be neither desirable nor undesirable. It was established solely to have some inherent meaning for benchmarking. The prime characteristic of importance for interpreting the scale is to recognise that the absolute productivity values are measured on a ratio scale with an absolute zero. That is, for example, an institution with a score of 0.25 can be thought of as half as productive as an institution with a score of 0.5. Bear in mind that estimates reflect only the data collected and that the value weight ranges are used to signify thresholds encapsulating a range of plausible interpretations.

Figure 32 illustrates the resulting institutional trends when the absolute productivity figures from the initial time point in the period are mapped to the TFP change results over the period from 2007 to 2016. The trends in the figure reveal an important context for system-wide behaviour. They highlight the nuanced and sometimes counter-intuitive nature of interpreting productivity measurement results. The converging relative absolute TFP trends from Figure 32 show precisely the opposite phenomena as the diverging normative TFP change trends exhibited earlier in the chapter. The trends, however, do not contradict one another. Rather, they provide immense explanatory value.

One way to reconcile the illustrations of divergent trends from the analyses in the beginning of the chapter with the convergent trends of the current analysis is to recognise that education productivity is underpinning absolute performance and research productivity is driving change. That is, emphasising education output gives a more favourable portrayal of institutions' proximity to the system production frontier, but emphasising research output gives a more favourable portrayal of institutions' performance trajectory. Further, given the convergent nature of the absolute productivity thresholds, it appears that the system's institutions are homogenising with respect to their education and research input-output performance. The positive changes in research-emphasised productivity appear to be closing gaps between research and education performance and closing gaps with respect to plausible interpretations of absolute productivity.

Finally, to bolster conclusions made from the separate SFP change analysis above, research-emphasised TFP growth frequently appears to occur at the expense of education-emphasised TFP growth. And the single outlier institution—Murdoch University—with more positive education-emphasised trajectories, follows the same general rule, in that the positive education trajectory appears to be at the expense of the research trajectory. This finding further concerning opportunity costs associated with innovations to research and education production functions. In summary, the findings from the absolute productivity trends analysis serve to recalibrate productivity interpretations and to contextualise the findings from the productivity change analyses above.

Figure 32: Absolute TFP approximation and relative change estimates



## 8.5 Emergent Patterns and Deeper Dynamics

### 8.5.1 Emergent Patterns

The Group of Eight (Go8) institutions, with the exception of The Australian National University, show noticeably different productivity change trends than the rest of the field. Their thresholds for plausible interpretations of productivity are much narrower, and their productivity levels are consistent and stable across the period. This could be evidence that the Go8 has reached some sort of barrier in productivity improvement. On the other hand, it could speak to the limitations of the metrics to capture elements of value that are unique to the country's top tier of institutions. In the latter case, one could argue that they have reached the pinnacle of higher education provision in the country, and because of these elite institutions' high quality of inputs in terms of staff and students, they are free to invest in broader, less measurable aspect of performance. Or, is it important to take the trends at face value and surmise that there is significant room for improvement in these institutions? Is it the case that because of their historical success, they have not had the burning platform or appropriate incentives make improvements to their productivity?

Three other universities in the dataset showed unambiguous sustained growth in both academic functions over the period. Those universities are Australian National University (ANU), Victoria University (VU), and Swinburne University of Technology (SUT). Each of these three institutions, however, started from the lowest base in terms of absolute productivity at the beginning of the period. The results suggest that, as compared to the rest of the field, there may have been more obvious inefficiencies to eliminate than at other institutions. Institutional case studies could shed light on whether measured trends at particular institutions may be attributable to unique management strategies, leadership styles, policies or organisational restructuring.

At least 11 institutions stood out for how they closed a gap between education and research emphasised absolute productivity portrayals. The institutions include Australian Catholic University, Central Queensland University, Curtin University of Technology, Griffith University, Macquarie University, Queensland University of Technology, The University of Notre Dame, University of Southern Queensland, University of Technology Sydney, University of Western Sydney, and University of the Sunshine Coast. Each institution achieved sustained productivity growth in research at the expense of growth in education—and in many cases decreases in education productivity. Further, many of these universities were originally established as non-university tertiary institutions, such as Colleges of Advanced Education (CAE), Institutes of Technology or as Colleges of Technical and Further education (TAFE). This suggests that they began with missions more focused on education than research. It is possible that, as universities pursuing heavier research objectives, education productivity has remained stagnant or has decreased because of shifting priorities in performance improvement.

Four institutions are worth noting as relative outliers in the field. First, SUT maintained one of the highest TFP change indexes among all institutions in the dataset. However, even

though most institutions' SFP productivity estimates were higher than TFP, SUT had one of the lowest SFP change indexes in the field. This likely indicates a significant operational and contextual difference between Swinburne University and the rest of the institutions in the country. The fact that SUT partnered with Online Education Services (OES), a private online education provider, in 2011 to administer its online course programs may explain the difference. Murdoch University is the only university in the country to have consistently increased education productivity at the expense of research productivity. It is the only institution in the field with a consistently diverging absolute productivity threshold. Third, Australian Catholic University's absolute TFP is higher than any other institution's in the field. While its education-weighted absolute TFP did not increase over the period, its research-weighted absolute TFP grew significantly to match its education productivity. It stands in contrast to Charles Sturt University. Charles Sturt started the period similarly in terms of its absolute TFP threshold with a very high level of estimated education productivity, but both of its academic functions declined in productivity consistently over the period. Fourth, Charles Darwin University is the only other university outside of the Go8 to show a very narrow threshold for productivity interpretation. Its base level of absolute productivity, however, is the lowest in the field. This finding suggests operational characteristics different to any other university in the country.

## 8.5.2 Dynamics of Data Elements

To gain a sense of what is driving productivity change across Australian institutions, individual component indexes are examined in conjunction with the composite productivity index. The analysis focuses on system-wide trends and how the drivers of productivity change often differ across different institutional contexts. Several of the data element component indexes listed in the introduction section in Table 39 are emphasised.

To contextualise major trends and their drivers, first consider that TFP change with equal education and research value weights across the period was measured at 0.984. This signals a slight drop in productivity by approximately 1.6%. Productivity is estimated to have peaked in 2013 at almost 2% over 2007 levels. The system is then shown to have experienced slight, albeit steady declines since that point. The slight decline in productivity is explained by the relative growths of input and output indicators. Inputs have grown to levels just under 150% of what they were in 2007. Research output has grown to over 155% of 2007 levels. Education output, however, has grown to about 135% of its level at the beginning of the period, signalling a rate of growth that has not kept pace with that of inflation-adjusted inputs. Details are provided in Table 39

**Table 39: System Drivers of 50-50 TFP and SFP Change 2007-2016**

| TFP Input | $X_{i\Delta t}^{TF}$ | $X_{\Delta t}^{TF}$ | SFP Input | $X_{i\Delta t}^{SF}$ | $X_{\Delta t}^{SF}$ | Edu Output  | $E_{i\Delta t}$ | $E_{\Delta t}$ | Res Output | $R_{i\Delta t}$ | $R_{\Delta t}$ | $Y_{\Delta t}$ | $P_{\Delta t}^{TF}$ | $P_{\Delta t}^{SF}$ |
|-----------|----------------------|---------------------|-----------|----------------------|---------------------|-------------|-----------------|----------------|------------|-----------------|----------------|----------------|---------------------|---------------------|
| $L_1$     | 1.41                 | 1.47                | $L_3$     | 1.77                 | 1.30                | $E_1$       | 1.37            | 1.35           | $R_1$      | 1.34            | 1.55           | 1.44           | 0.98                | 1.11                |
| $L_2$     | 1.48                 |                     | $L_4$     | 1.32                 |                     |             |                 |                | $R_2$      | 1.40            |                |                |                     |                     |
| $I$       | 1.49                 |                     | $L_5$     | 1.08                 |                     | $R_3 + R_4$ | 2.29            |                |            |                 |                |                |                     |                     |
| $K$       | 1.52                 |                     | $L_6$     | 1.33                 |                     |             |                 |                |            |                 |                |                |                     |                     |

The growth rates of individual input components remained largely consistent with one another across the period. Capital services grew slightly more quickly than the others, and academic labour expenditures grew least quickly. FTE staff numbers, however, exhibited unique trends, with teaching only staff have grown by over 75% and traditional teaching and research staff having grown by only 8%. For research outputs, adjusted scholarly output grew at a pace that more than doubled the rate of any other output or input factor, giving a sense that institutions across the system have prioritised and likely incentivised the production and dissemination of scholarly works over other university functions. For education, the growth of equivalent full-time load has outpaced the growth rate of completions. This is not surprising. There is an inherent delay between student enrolment and completion. If figures for EFTSL growth are examined three years prior to end of the period, the cumulative change result is estimated at 1.283, which is nearly identical to the completion growth rate by 2016. The analysis estimates that the real rates of growth for student load and completions are in lockstep with one another, signalling that undergraduate progression rates have likely remained constant as the system has grown.

There is no single story for the drivers of productivity for individual institutions. While Australian universities are scaling up across the board, evidence suggests that individual institutions have different experiences with success at scale. Table 40 shows the TFP component indexes for four selected universities to illustrate the magnitude of differences. Consider Swinburne University of Technology and The University of Sydney. The institutions each increased their monetary inputs by similar figures, 47% and 46% respectively. Swinburne, however, increased its education outputs and research outputs by 88% and 82%, respectively, while Sydney increased its education and research outputs by only 28% and 26%.

**Table 40: Diversity of institutional TFP growth trends 2007-2016**

| Institution                        | $X_{\Delta t}$ | $E_{\Delta t}$ | $R_{\Delta t}$ | $P_{\Delta t}^{TF}$ | Rank |
|------------------------------------|----------------|----------------|----------------|---------------------|------|
| Swinburne University of Technology | 1.468          | 1.881          | 1.817          | 1.259               | 4    |
| Central Queensland University      | 1.008          | 0.812          | 1.711          | 1.170               | 7    |
| Murdoch University                 | 1.240          | 1.588          | 0.973          | 1.003               | 22   |
| The University of Sydney           | 1.457          | 1.276          | 1.264          | 0.872               | 35   |

Different behaviours in component indexes likely indicate different practices in resource management, operations, and policies for growth. While broad patterns in education and research may show trends of convergence for institutional input and output performance, deeper analysis shows how institutions remain diverse. Decomposing the productivity metrics suggests specific and unique strategies, as well as the potential effectiveness of the strategies. The evidence above can give rise to further targeted investigation. Which institutions drive productivity by limiting inputs and focusing on efficiency? Which drive productivity by growing outputs and focusing on value creation? Which have found gains by shifting production focus from one output type to another? Finally, based upon the field of institutions, which strategies in context appear most sustainable, desirable, and transferable? To maintain focus on only the final productivity indicator, or to stick to descriptive

institutional portrayals of distance to a theoretical production frontier masks both the urgency of these questions and the foundations of evidence for addressing them.

Different patterns and behaviours across institutions are summarised in Table 41. The table explores three types of productivity measured above for each institutions: joint TFP for education and research, education SFP and research SFP. Labels are given to the type of change that occurred for each type of productivity for each institutions, based on the classifications provided in Box 2 and Box 3 from Chapter 3. Because Australian universities experienced consistent growth over the period, the two common types of change include ‘Scaling up benefits’ and ‘Net expense growth’. Respectively, they refer to output growth increasing more quickly than input growth and *vice versa*. ‘Improving effectiveness’ was also common, referring to quick growth in outputs and static or little change in inputs.

**Table 41: Classifying productivity change trends**

| <b>Institution</b>                     | <b>Joint TFP</b>        | <b>Education SFP</b>    | <b>Research SFP</b>     |
|----------------------------------------|-------------------------|-------------------------|-------------------------|
| Australian Catholic University         | Scaling up benefits     | Scaling up benefits     | Improving effectiveness |
| Australian National University         | Scaling up benefits     | Scaling up benefits     | Full Innovation         |
| Central Queensland University          | Improving effectiveness | Full decline            | Full innovation         |
| Charles Darwin University              | Scaling up benefits     | Scaling up benefits     | Scaling up benefits     |
| Charles Sturt University               | Net expense growth      | Scaling up benefits     | Scaling up benefits     |
| Curtin University of Technology        | Net expense growth      | Neutral growth          | Scaling up benefits     |
| Deakin University                      | Scaling up benefits     | Scaling up benefits     | Improving effectiveness |
| Edith Cowan University                 | Scaling up benefits     | Improving effectiveness | Improving effectiveness |
| Federation University                  | Net expense growth      | Declining efficiency    | Scaling up benefits     |
| Flinders University                    | Net expense growth      | Scaling up benefits     | Scaling up benefits     |
| Griffith University                    | Scaling up benefits     | Scaling up benefits     | Scaling up benefits     |
| James Cook University                  | Net expense growth      | Scaling up benefits     | Scaling up benefits     |
| La Trobe University                    | Scaling up benefits     | Improving effectiveness | Improving effectiveness |
| Macquarie University                   | Net expense growth      | Scaling up benefits     | Scaling up benefits     |
| Monash University                      | Scaling up benefits     | Improving effectiveness | Improving effectiveness |
| Murdoch University                     | Scaling up benefits     | Scaling up benefits     | Full decline            |
| Queensland University of Technology    | Net expense growth      | Scaling up benefits     | Scaling up benefits     |
| RMIT University                        | Scaling up benefits     | Scaling up benefits     | Scaling up benefits     |
| Southern Cross University              | Net expense growth      | Declining efficiency    | Improving effectiveness |
| Swinburne University of Technology     | Scaling up benefits     | Scaling up benefits     | Scaling up benefits     |
| The University of Melbourne            | Net expense growth      | Scaling up benefits     | Scaling up benefits     |
| The University of New England          | Scaling up benefits     | Improving effectiveness | Improving effectiveness |
| The University of New South Wales      | Net expense growth      | Scaling up benefits     | Scaling up benefits     |
| The University of Newcastle            | Net expense growth      | Scaling up benefits     | Scaling up benefits     |
| The University of Notre Dame Australia | Scaling up benefits     | Scaling up benefits     | Improving effectiveness |
| The University of                      | Net expense growth      | Scaling up benefits     | Scaling up benefits     |

|                                     |                         |                         |                         |
|-------------------------------------|-------------------------|-------------------------|-------------------------|
| Queensland                          |                         |                         |                         |
| The University of Sydney            | Net expense growth      | Scaling up benefits     | Scaling up benefits     |
| The University of Western Australia | Scaling up benefits     | Improving effectiveness | Scaling up benefits     |
| University of Adelaide              | Net expense growth      | Scaling up benefits     | Scaling up benefits     |
| University of Canberra              | Scaling up benefits     | Improving effectiveness | Improving effectiveness |
| University of South Australia       | Net expense growth      | Comprehensive decline   | Scaling up benefits     |
| University of Southern Queensland   | Scaling up benefits     | Declining efficiency    | Improving effectiveness |
| University of Tasmania              | Net expense growth      | Scaling up benefits     | Scaling up benefits     |
| University of Technology, Sydney    | Scaling up benefits     | Scaling up benefits     | Improving effectiveness |
| University of the Sunshine Coast    | Scaling up benefits     | Scaling up benefits     | Improving effectiveness |
| University of Western Sydney        | Scaling up benefits     | Scaling up benefits     | Improving effectiveness |
| University of Wollongong            | Scaling up benefits     | Scaling up benefits     | Improving effectiveness |
| Victoria University                 | Improving effectiveness | Scaling up benefits     | Improving effectiveness |

A notable finding from Table 41 is that only four institutions out of 38 have the same change profile across three productivity indicators. The different observations of behaviour across different inputs and outputs likely signify different modes of operation and execution of university functions. Opposing trends, such as those of Central Queensland University and Murdoch University, signify likely trade-offs associated with prioritising the productivity of one function over another. Aligned trends, such as those of RMIT and Swinburn, may signify mutually reinforcing operations and strategy across functions. No combination of types of change across the indicators is inherently desirable or detrimental for an institution. Depending upon institutional mission, objectives, priorities, and other unmeasured factors, one type of productivity change result might be viewed favourably for one institution but negatively for another. The results highlight the importance of using multiple productivity metrics for any productivity assessment, as well as the importance of using measured trends as a baseline for conducting further investigation that can give the metrics more meaning.

### 8.5.3 Model Versions Insights

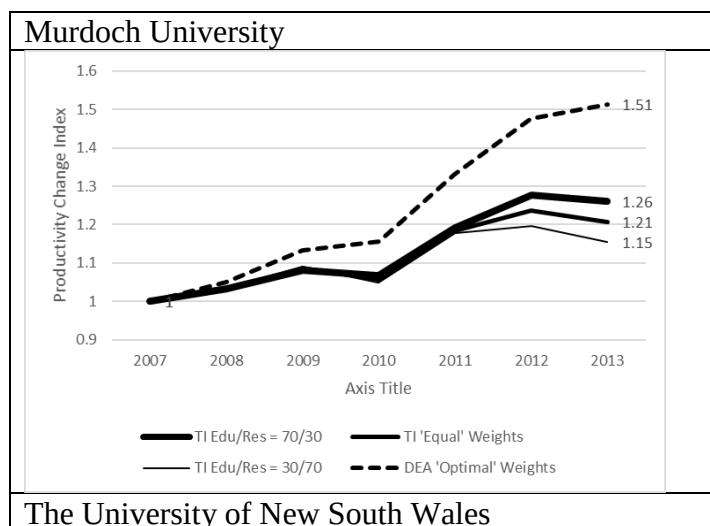
This section briefly covers key findings from comparing the results of multiple versions of the model in this research. **Appendix I** provides several tables and all the details from which the following conclusions were made. First, while TFP productivity change results did vary across versions of the model, when aggregated to the system level, the results of the models did not vary by more than a few percentage points. The system TFP change scores for the first version and the final version were most similar. Before adjusting for inflation, from 2007 to 2013, the equally weighted output cumulative TFP change index at the system level for the first version of the model is 0.887. The index for the final model is 0.883. This speaks to the reliability of the findings and the robustness of, especially, the final version of the model, which uses a larger number of data elements to generate estimates.

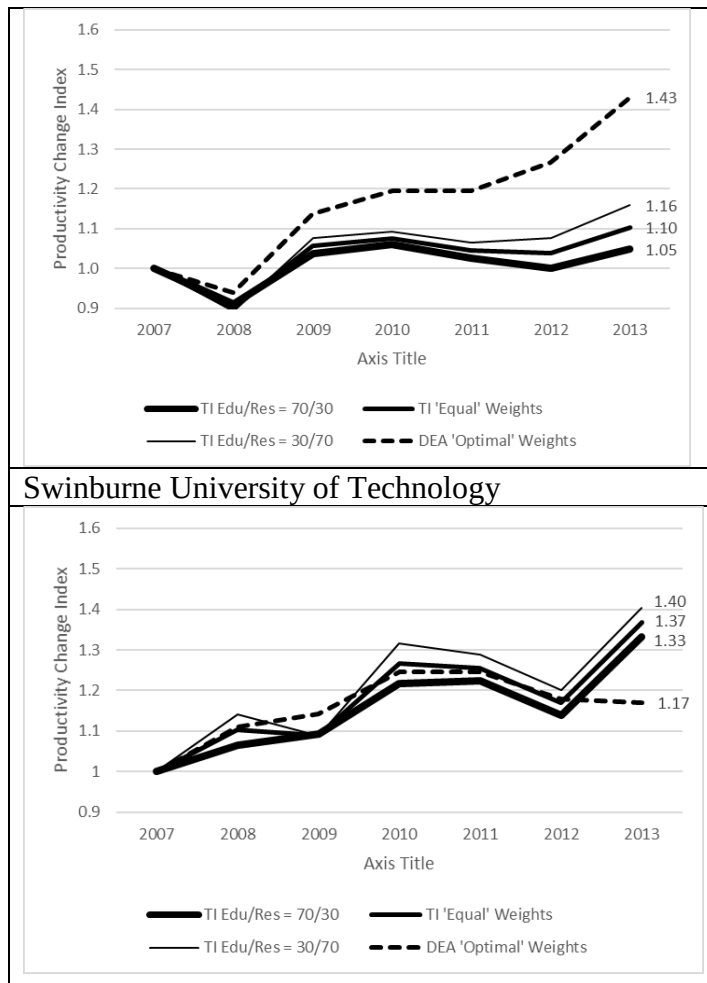
The correlation between different input and output indicators, which allowed for similar result across versions of the model led another notable finding. At the institution level, TFP

indexes between model versions that used different data elements, but which used consistent output value weighting schemes differed less than TFP indexes for a single model version using different output value weighting schemes. Assuming sufficient accuracy of inputs and outputs, this is not a trivial finding. It suggests that there may be enough co-variance between respective institutional inputs and institutional outputs that treatment of the data in a measurement model is more consequential than what data is selected for measurement. This further supports the idea that measurement techniques need not only to be developed with utmost care and for specific purposes, but also that they should be as transparent as possible.

With respect to adjusting for the real value of money, two findings were taken into consideration. Australian CPI data, show that from 2007 to 2016, currency inflation rose by approximately 20%. This figure by itself is enough to justify adjusting figures in the productivity calculation so that price fluctuations are not over-represented in the productivity calculation. Further, after adjusting for the real value of money, results appeared more positive, and trended more in the direction of what previous studies on the topic have shown. Even when adjusting for currency inflation, however, results from the model in this research during the period 2007-2013 were still significantly different from those of prior studies. Paired with the finding from the paragraph above, it sparked curiosity about replicating the Moradi-Motlagh et al. (2016) DEA study using the same exact data elements they used in their model, but instead aggregating the inputs and outputs with TI techniques. The input-output framework from the DEA study is one input and two outputs. The input is total annual operational expenditures. The education output is total equivalent full-time student load, and the research output is total adjusted publications. Moore et al. (2019) show the results of replicating the study with a TI model, where the annual expenditures are adjusted for currency inflation. Results are presented in Figure 33.

**Figure 33: Comparison of productivity TI and DEA estimates for three select universities (Moore, Croucher, & Coates, 2019)**





The exercise highlights the indeterminacy of productivity analysis when using different treatments of data and different aggregation methods. For Murdoch University it is apparent that the DEA algorithm emphasised education outputs to a greater extent than any specification of the TI, and for The University of New South Wales, the DEA algorithm further emphasised research. For Swinburne, attempting to reverse engineer or understand how the DEA algorithm determined weights is more difficult. These findings further highlight the importance of transparency in measurement if results are to be used for practical purposes.

## **8.6 Summary**

### **8.6.1 Focus and Contributions**

This chapter has provided all quantitative results on Australian universities from the productivity measurement model in this research. The chapter further analysed and interpreted the results provided. Using the findings as evidence, it has established the foundations for a productivity and performance assessment of Australian universities and has demonstrated an expanded scope for insight that productivity measures can deliver as performance indicators.

Several key findings are worth reiterating. First, with respect to change in TFP, positive institutional trends typically depend on calculations that emphasise performance in research. That is, growth in standard research output measures exceeded the rate of growth in standard education output measures from 2007 to 2016. The tendency for research productivity change to outpace that of education productivity was also demonstrated with separate estimates of productivity change for the two academic functions. Together, TFP and separate productivity change estimates appear to demonstrate measurable trade-offs associated with improving the productivity of one function over the other. There were three notable exceptions where institutional productivity change of both academic functions increased consistently over the period. When relative change results were estimated using absolute productivity approximations, however, each of these three outlier universities exhibited significantly lower absolute productivity levels at the beginning of the period than the rest of the universities in the field. That is, sustained growth in both academic functions appears to be highly associated with a need or a drive to achieve productivity levels that are on par with peer institutions.

The relative productivity change calculations also cemented conclusions that there exist performance trade-offs associated with education and research. Levels of education productivity at the beginning of the period were portrayed as more positive than those of research productivity, but by the end of the period, levels of research productivity had risen. With this behaviour observable across the system, findings also appear to support a homogenisation of performance across institutions, at least in the dimension of productivity. Finally, results also showed more positive rates SFP change than TFP change. The relatively quicker growth rates of labour productivity could have many explanations, but findings from the current research are not sufficient to offer an explanation. This is an area where further research could add insight.

### **8.6.2 Considerations and Limitations Arising**

The quantitative findings presented in this chapter disclosed a host of productivity and performance dynamics observable across institutions within the Australian higher education system. Measured results, however, tell only part of the story of performance. As discussed in the previous chapter, the 20 key-informant interviews provided substantive qualitative evidence for performance trends within the system. A synthesis of the qualitative and

quantitative information has great potential to inform policy, decision-making and the execution of new productivity and performance initiatives. In the final chapter of this dissertation, quantitative results will be squared with qualitative findings from Chapter 7, as well as with additional qualitative evidence from the interviews not yet presented. A synthesis of all findings from this research exposes implications for international practice in university productivity measurement, as well as implications for policy and provision of higher education.

## CHAPTER NINE: DISCUSSION OF FINDINGS AND IMPLICATIONS

*I think we are ripe for some big disruptive change, and if we don't change ourselves, we will be changed from the outside...*

*...it's undiscussed undiscussable problems that cause the biggest issues because you can't find solutions if you can't be in a position to admit a problem.*

–Participant KS.9  
Academic Program Director

### **Abstract:**

This chapter describes the contributions of this PhD project to knowledge, theory and practice. Learnings from the literature review and the design science research stages are synthesised and triangulated to form conclusions about higher education productivity and its measurement in Australia. The chapter first reviews findings from previous chapters and then uses evidence to address each research question. Design principles for productivity measurement models are discussed, along with broader insights about measurement results. The next section discusses implications for policy and practice and explains how research findings inform three current policy issues in Australia: performance funding, provider category standards, and financial transparency. Limitations of the research are then discussed. They include the exclusion of quality indicators, service and engagement, and inter-departmental university dynamics. The chapter concludes by urging further research in targeted areas and reiterates key domains where research findings have immediate value and utility.

### **Key Words:**

Productivity paradigms, policy, practice, efficiency, value creation

## 9.1 Research Summary

### 9.1.1 The Value of the Research

This chapter covers research Stage 8, synthesis and reporting of all findings. The dissertation has demonstrated the value of measuring productivity change using Törnqvist indexes. The research has implications for global practice in using productivity indicators for university performance assessment and has specific implications for the provision of higher education in Australia. The study's literature review discussed rationales for measuring and assessing university performance and characterised diverse stakeholder interests in higher education performance. The review described emerging and evolving operational challenges and identified productivity as a key indicator of institutional performance. Research following the literature review investigated alternative methods for measuring productivity in higher education. It has demonstrated the value of conducting research to refine measurement methods and approaches and now synthesises qualitative and quantitative findings.

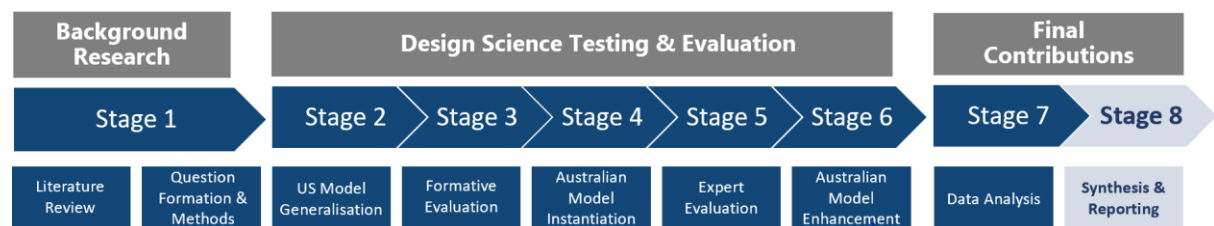


**Figure 34: Research Stage 8**

The strengths of indexing methods are distinct from those of other common measurement techniques. While common, black box measurement methods allow for precision in generating statistically robust estimates in relation to a dataset, easily decomposed indexing methods allow for precision in tailoring quantitative techniques to fit broader measurement contexts. The value of index measures is greater flexibility in translating stakeholder performance concerns into quantitative information. Early findings indicated that measuring productivity change—rather than absolute institutional productivity—would be most useful. Further, the multi-stage design process—involving higher education practitioners, stakeholders and experts—has emphasised how key learnings are derived from the exercise of conducting measurement and from refining measurement techniques. The development of a fit-for-purpose measurement tool, however, has also allowed for a unique university performance assessment of Australian universities.

### 9.1.2 Accumulated Learnings

Practitioners across nine countries in Asia-Pacific provided initial insight about the fundamental strengths and limitations of the TI aggregation method. Each country context exhibited unique limitations and availability of data, as well as varied interests in

measurement results. Prime results from the international test of the model include (A) a validation of TI methods for use in higher education performance assessment and (B) a refining in scope of the model for Australian higher education. On one hand, the initial model both informed and reflected higher education contexts in each country. On the other, it was determined that model refinement for the Australian context should be more discriminatory for the types of data considered. A more limited model can produce more reliable, targeted results—albeit at the expense of incorporating a broader set of interests in performance.

Subsequent stages of research revealed specific performance interests in Australian higher education. Key insights are two-fold. First, findings from the interview stage of the research guided the design process and pointed to further model adaptations best suited for the Australian context. Second, broader insight from the interviews suggests how international practice in university productivity measurement could be improved for more reliable performance assessment. Leveraging the model to generate multiple portrayals of productivity trends serves as a prime example. Portrayals of SFP vs. TFP, joint vs. separate, and normative vs. relative change speak to different interests in performance, aspects of performance and different processes involved with academic work and university operations. Exposing both complementary and contrasting trends within the data provides a more nuanced and complete picture of performance. Further, successive implementations of subsequent versions of the model—and the variation in model version results—illustrates the care that must be taken in using quantitative indicators in performance assessment. One on hand, the exercise reinforces a healthy scepticism in the available data. On the other, successive iterations of the model unpacked and characterised different aspects of complexity and contribute to a roadmap for keener use and interpretation of productivity data for performance assessment.

Final stages of research provided evidence for a performance assessment of Australian universities. Findings describe both cross-cutting institutional trends and fringe cases. When measurement results are squared with interview responses, a larger and more reliable story unfolds. The quantitative and qualitative evidence has implications for policy and practice in Australian higher education. Implications concern the legislation of public policy and arbitration and decision-making among institutional leaders and key stakeholders. The remainder of this chapter discusses the final synthesis of research findings and makes conclusions based on evidence generated from all stages of research.

## **9.2 Findings and Implications**

### **9.2.1 Presentation of Findings**

Findings are organised and presented to address this study's research questions. Findings thus speak to theoretical aspects of higher education productivity measurement, then to technical issues for conducting measurement, and finally to observed trends that could be useful in performance assessment and decision-making. For reference, the questions are restated below:

**Q1:** *How can different approaches to understanding higher education productivity be characterised?*

**Q2:** *What characteristics of a productivity measurement model are appropriate for investigating Australian universities?*

**Q3:** *How can the productivity of Australian universities be characterised?*

## 9.2.2 Addressing Research Question 1

### 9.2.2|a. Approaches to understandings

Three broad approaches to understanding productivity are identified in this research. The approaches are coined, ‘methodological’, ‘archetypal’ and ‘functional’. They are not mutually exclusive. Rather, they each play a role in any effort to further understand productivity. Findings from this research—as well as prior literature—illustrate how one approach may be dominant under particular circumstances and contexts. Figure 35 summarises how each approach may form the baseline for enquiring about the subject.

The methodological approach seeks to understand productivity by perfecting the processes and techniques required to generate productivity estimates for institutions. The archetypal approach focuses on different notions and interpretations of higher education productivity as a construct. The functional approach seeks to understand through the examination of available data and focuses on the nature, strengths and limitations of what information is at hand. The following paragraphs further explain each approach.

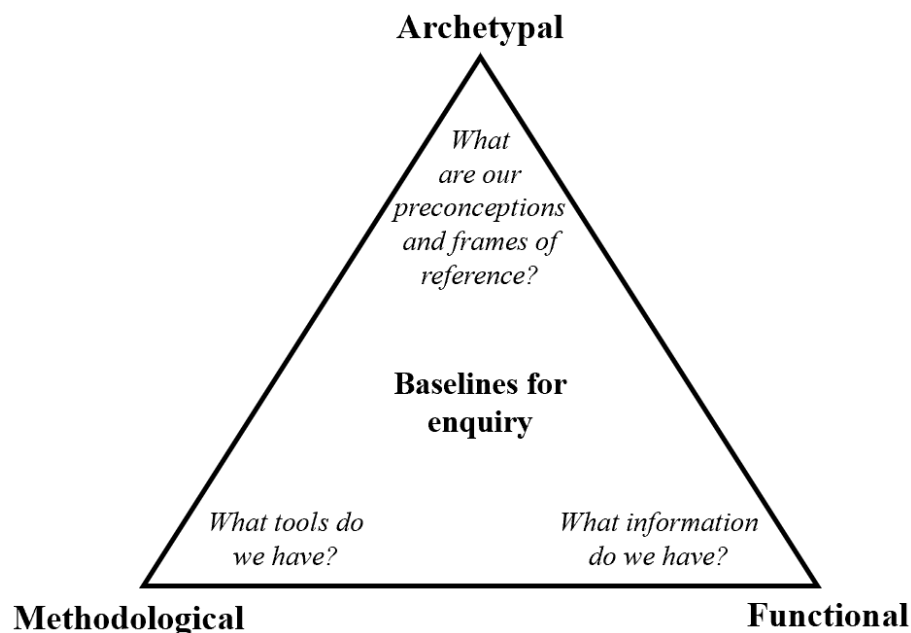


**Figure 35: Approaches to understanding higher education productivity**

### 9.2.2|a. Methodological approaches

Methodological choices and framings affect the nature and the utility of results. A distinguishing feature of the current study is its use of indexing methods. The approach has been contrasted to more common linear programming techniques used to estimate university productivity. Approaches to measurement are not innocuous. Technical conventions can

serve to reinforce overarching epistemological approaches, and thus reinforce the types of results that a study can produce.

Findings bring to light the difference between inductive and deductive approaches to understanding higher education productivity. Trochim, Donnelly and Arora (2015) summarise inductive approaches to research as prioritising the formation of theory via emergent patterns among documented observations. The deductive approach, by contrast, begins with theory on a topic of interest and seeks to confirm or refute the theory based on collected observations. Inductive approaches are more open-ended and exploratory, while deductive approaches are narrower and more precise (Trochim et al., 2015). The current research has sought to build understanding of productivity in higher education in an inductive manner, through the identification of patterns in collected data. As discussed in Chapter 3, however, most contemporary research featuring university productivity measurement upholds conventional economic assumptions about institutions and seeks to characterise productivity in relation to a production frontier. Most contemporary, deductive university productivity research is best positioned to offer high levels of precision in confirming individual institutions' proximity or trajectory with respect to an estimated frontier.

Deductive approaches to understanding productivity not only represent the dominant frame in higher education, but also that dominance is reinforced by measurement convention. Participant interview responses illustrate this point. Participant responses largely diverged from conventional academic framings of university productivity. Across 20 key stakeholder and technical expert interviews, no participant defined productivity with reference to a production frontier, nor framed any productivity issue with respect to relative technical efficiencies between institutions. Further, no participant inquired about the capacity of the measurement model in this research to generate efficiency frontier approximations.

This dissertation has acknowledged the distinction between deductive and inductive methodological approaches, as well as the relative prevalence of deductive methods. As articulated by De Witte & Sneyers (2017), most efficiency measurement research in education begins with assumptions about the suitability of linear programming methods. Analysis that follows subsequently focuses on fit, reliability and precision with respect to a particular dataset (De Witte & Sneyers, 2017). Opportunities for broader insight are likely being missed with such a focus. The inductive approach in this dissertation has demonstrated broader compatibility with a diverse set of interests and has offered technical alternatives to represent those interests with quantitative data.

### **9.2.2|b. Archetypal approaches**

Archetypes and preconceptions about productivity as a construct also influence approaches to understanding. The productivity paradigms identified in Chapter 7 serve as the prime example. They represent individual conceptualisations of the phenomenon based on past experiences and performance interests. They include 'value creation productivity', 'efficiency productivity', 'process-oriented productivity' and 'pragmatic productivity'. The significance of their identification and explanation is two-fold. First, they serve as the first empirically-grounded framing of broad productivity conceptualisations specific to the higher education sector. Second, these paradigms do not arise from a review of technical definitions identified in scholarly articles, as with Ghobadian and Husband (1990) and Pritchard (1995). Rather, using the literature as an anchor point, the paradigms in this research emerged from an

analysis of productivity definitions provided by key higher education stakeholders and experts not associated with the academic study of productivity or production theory. As acknowledged in Chapter 7, the sample of participants was never intended to be representative of a population, so classifications and summaries of these paradigms could change with a larger sample of participants. The paradigms' alignment with conceptual elements already identified in the literature—and with the emergence of a new conceptualisation—however, indicates the significance of the findings.

While the paradigms are intended to represent mutually exclusive conceptualisations of productivity, they are not exclusive to use by individuals. The interview phase of this research demonstrated that higher education stakeholders often took more than one perspective on productivity. Most interviewees did, however, exhibit a proclivity to adhere to one of the paradigms. The principle implication of this finding is that no paradigm is more valid than another. Each perspective on productivity spawned unique conversation about real issues and interests in the sector.

The paradigms have strengths and limitations. Acknowledging them may help to facilitate more efficacious arbitration and decision-making among stakeholders. Evidence from the interviews suggests, for example, that the value creation productivity paradigm often clashes with the efficiency productivity paradigm. Objectives of financial prudence unsurprisingly complete with objectives to maximise value creation. The two framings need not inherently complete, however, as is evident in conversations with stakeholders who exhibited pragmatic framings of productivity. Using this frame—and with an understanding of limitations associated with scarce resources—objectives to maximise outputs and minimise inputs may exist simultaneously. Pragmatic productivity framings, however, may clash with process-oriented framings in terms of what interventions are viewed to support or facilitate productivity improvement. One participant noted,

*I think it's a laudable thing to be as productive as possible. But I think sometimes the very emphasis on productivity makes it harder to be productive.*

—KS.9

### **9.2.2|c. Functional approaches**

Some approaches to understanding higher education productivity are functional and utilitarian. The literature review from Chapters 2 and 3, as well as findings from Chapter 6, demonstrate how a model's input-output framework embodies an explicit approach to understanding productivity. The framework represents operative interests and values of stakeholders. The interests may be those of analysts or those who committed resources to collecting and disseminating source data used in the analysis. The pragmatic choices that APO practitioners made to either include or exclude certain indicators, for example, reflect not only the state of their respective countries' information systems, but also reflect larger higher education interests present in each country. The myriad different data combinations that could be chosen for an input-output framework reinforce alternative understandings of higher education productivity, and the source data available is no coincidence with respect to the interests of those who make it available.

Input-output framings represent baseline assumptions about what constitutes a valid interpretation and understanding of productivity. Nazarko (2014), for example, considers

student achievements as educational outputs by including numbers of scholarship recipients and study abroad participants. Olariu & Brad (2017), however, count raw enrolment numbers and government revenue. And although the types of university performance being represented are different, each study presents results under the guise of ‘productivity’.

A model’s input-output framework is its key transparent artefact to help stakeholders understand the operative interests and values with a single interpretation of productivity. The development and evolution of the Australian model in this research speaks to several trade-offs for framing the productivity of the system. Model development involved addressing multiple pragmatic questions, such as whether or not to (A) distinguish between output and outcome indicators, (B) to incorporate value in conjunction with input and output quantities and (C) whether to choose indicators valued more by some stakeholders than by others. Each decision represents a pragmatic choice that ultimately defines an approach to understanding higher education productivity. Any input-output framework will always be more or less appropriate for embodying stakeholder values and interests, and any framework should be chosen deliberately to fit a predetermined and stated purpose for the measurement exercise.

### 9.2.3 Addressing Research Question 2

#### 9.2.3|a. Principles of design

Six inherent characteristics of the final measurement model underpin its technical suitability for investigating higher education productivity. Table 42 presents the characteristics as replicable and testable principles of design. They are intended to help inform future modelling and design decisions by highlighting key features and technical specifications that can be refined and manipulated. The design principles are presented as overall learnings and reference specific applications from the current model. As discussed in Chapter 4, Plomp and Nienke (2007) argue that—after the artefact itself—the prime output of design science research should be the principles that underpin the artefact’s design. The design principles should reflect learnings from both the scientific literature and the practical needs and concerns that the model was designed to address within its operating context.

**Table 42: Productivity measurement model design principles**

| Principle of Design                         | Nature of Principle            | Technical Benefit    |
|---------------------------------------------|--------------------------------|----------------------|
| 1. Institutional Comparability              | <i>Intentionally Developed</i> | External Validity    |
| 2. Multi-year Input-Output Linkages         | <i>Discovered</i>              | External Validity    |
| 3. Output Value Accounting                  | <i>Intentionally Developed</i> | External Validity    |
| 4. Alternative Productivity Representations | <i>Discovered</i>              | Construct Validity   |
| 5. Model Component Consistency              | <i>Discovered</i>              | Internal Consistency |

|                         |                   |                      |
|-------------------------|-------------------|----------------------|
| 6. Experimental Control | <i>Discovered</i> | Internal<br>Validity |
|-------------------------|-------------------|----------------------|

The third column of Table 42 indicates how each design principle is tied to scientific measurement standards. The general extent of the final measurement model's grounding both scholarly knowledge and the Australian higher education environment has been demonstrated by the literature review and design science process of this research. Framing the design principles in terms of broader scientific measurement standards, however, is both helpful and necessary for future research. De Witte (2017) further emphasises how educational efficiency research lacks sufficient influence from broader fields of measurement.

McBurney and White (2010) identify and discuss four key scientific measurement and assessment standards. The principles are 'external validity', 'construct validity', 'internal validity', and 'internal consistency'. External validity refers to a model's ability to be applicable in external circumstances, not confined to the study or experiment. It speaks to the model's generalisability in producing results for broader contexts than that of the research environment. Construct validity refers to the extent to which the identifiable individual components of the model accurately represent the ideas and concepts that they are purported to represent. The authors note that it is impossible to ensure construct validity, but a research project can be planned so that it is more plausible—which is the objective of the design science model assessment criteria. Internal validity is the ability to produce results that are true and accurate within the context of the study. It speaks to the extent to which potential confounding and external variables are controlled for within the research environment. Internal consistency is more difficult to define with respect to a quantitative model, as even some efficiency studies equate internal validity with internal consistency (Tsionas, Malikov, & Kumbhakar, 2018). Differentiating internal consistency, however, is helpful with respect to the current research. Appelbaum (1978) refers to internal consistency of productivity measurement models as the quality of different model components being defined and related to one another in a consistent manner. Internal consistency is achieved with common and structured relational parameters for all model components. The follow sections define and describe each measurement model design principle uncovered in this research and explain their adherence to the four scientific measurement standards.

### ***9.2.3|b. Explanation of design principles***

The first design principle is **institutional comparability**. Its incorporation improves external validity by allowing for more generalisable results, and it represents an intentional design standard for the model. The prime unit of analysis in this research is the institution. Comparing measurement results across institutions despite their unique contextual orientations has been a concern from the beginning of the research. Three aspects of the model reflect institutional comparability:

1. **Data aggregation method.** The TI is a measure of productivity change that allows for comparisons of institutions on their own terms. Each institution's initial productivity indicator is normalised at a value of '1' at the beginning of the period examined. Any institution may thus directly compare its rate of change to that of another institution with no priors or value judgements concerning initial productivity levels.

2. **Data selection.** Sixteen of the eighteen data elements used in the final model are Australian university key performance indicators with consistent data collection and reporting requirements, as outlined by the DET. The remaining two indicators are collected and reported globally on a consistent basis as outlined by Elsevier Research Intelligence (Colledge & Verlinde, 2014).
3. **Absolute productivity approximation.** As discussed in Section 7.4.6, the TI indexes can distort rate of change portrayals when initial productivity values are normalised to a value of '1'. The effect may be to amplify or to dampen actual rates of change. The absolute productivity approximation provides an alternative perspective on productivity change rates by applying a range of plausible value judgements on initial institutional productivity levels per institution.

The second design principle is **multi-year input-output linkages**. Its incorporation improves external validity by representing input-output dynamics that are more characteristic to universities in general. This design principle was discovered during the first evaluation of the model and validated during the second evaluation. The design principle is reflected in the 'smoothing' of input data elements described in Section 6.4.4. Taking the arithmetic mean of three consecutive years of inputs in the current model could be improved upon with deeper investigation of individual relationships between inputs and outputs, but the overarching principle is established to help account for the longer-term nature of output production in universities that may not be captured reliably with direct year-on-year estimations.

The third design principle is **output value accounting**. Its incorporation improves external validity by aligning with broad interests in measuring more than raw quantities associated with student numbers or research churn. This design principle was intentionally developed to address stakeholder concerns raised in Chapter 2 and Chapter 7 about productivity conclusions being made with indiscriminate and less nuanced consideration of only output quantity dynamics. As described in Chapter 5 current methods for capturing the value of outputs are only a starting place. They add nuance to the estimates while still using standard units of measure, yet they do not represent the full value of higher education to individuals and society. Three aspects of the model reflect output value accounting:

1. **Adjusted load.** This indicator is described in Chapter 5. Its prime objective is to use an empirical foundation of evidence to systematically capture the additional value added when coursework students complete a degree.
2. **Adjusted load research.** This indicator uses the same logic and principles as adjusted load and applies the calculation to graduate research students.
3. **Adjusted publications.** This indicator is explained in chapter 7. It uses the same logic and principles as adjusted load, but it uses an alternative empirical base of evidence to make adjustments to publication quantities based on their field-weighted citation impact.

The fourth design principle is **alternative productivity representations**. Its incorporation improves construct validity through complimentary and comparative measures of key phenomena represented in the model. The model was intentionally designed for multi-input, multi-output productivity representation, but the current design principle is distinct from that intention, and the evidence supporting it was emergent. The first tests of the model, comparing the results from multiple versions of the model and results from prior studies

exhibited the difference in potential portrayals of productivity that can emerge from alternative framings of the concept. The range of explicit interests and productivity conceptualisations evident in the interviews reaffirmed the need to include alternate representations of key constructs in the model to provide a more complete and valid interpretation of observations and real-world phenomena. Three aspects of the model reflect alternative productivity representations:

1. **Multiple input and output construct indicators.** The productivity drivers analysis from Chapter 8 illustrates why multiple, alternative representations of input and output constructs is essential for a more valid interpretation of university productivity. The distinct data elements representing either inputs or outputs are sometimes co-variant but sometimes divergent. They allow for triangulation of dynamic information to generate a more valid interpretation of the constructs and their behaviour.
2. **Alternative input representations.** The SFP and TFP estimates—which include different input data elements—each tell different stories about Australian university productivity. While most institutions exhibited more positive SFP estimates, some did not, and having both estimates allows for more nuanced interpretation of overall trends.
3. **Alternative production complexity representations.** The joint vs. separate productivity measurements gave complimentary interpretations of productivity change within the system, and thus increase the validity of the estimates. But each also exhibited distinct elements and added depth to the interpretation of education and research productivity trends.

The fifth design principle is **model component consistency**. Its incorporation inherently improves the internal consistency of the model and further distances the measurement model from ‘black box’ calculation techniques. The incorporation of this design principle emerged from research findings. Section 7.4.7 explains the adaptations made to the model to embody this principle. They include calculating ‘adjusted load research’ and ‘adjusted publications’ to run consistent with the ‘adjusted load’ construct for education output. The consistent design of research and education output constructs facilitates more systematic accounting for what—and to what extent—individual output data elements are driving broader trends.

The sixth design principle is **experimental control**. Its incorporation increases the internal validity of the productivity calculation. A first step of adapting the NRC model in Chapter 5 was the specification of the non-priced output TI aggregation method, which requires researcher specified value weights for output components. The informative potential of testing value weight ranges, however, was an emergent finding. The ability to control for the emphasis and value weighting of data elements in the productivity calculation is an objective scientific advantage over automatically assigned value weights based on implicit characteristics of a dataset. The implications on final measurement results of different emphases on data elements are important in their own right for better understanding within-institution dynamins. Further, the magnitude of effect on final productivity estimates from different value weighting schemes is not consistent between institutions. Understanding which institutions are more sensitive to variation in data aggregation techniques is also helpful for interpretation of results.

### 9.2.4 Addressing Research Question 3

#### 9.2.4a. *The direction of change*

Directions or rates of growth and decline do not sufficiently characterise the nature of productivity change within the Australian system. Different empirical perspectives on system productivity show different results. A synthesis of multiple perspectives does, however, provide compelling evidence for some broad patterns. First, positive productivity change portrayals for most of the 38 universities in the system depend upon emphasising research output growth over education output growth. On the other hand, education-emphasised portrayals give institutions a more positive absolute productivity portrayal, effectively positioning them within closer proximity to the system's production frontier.

SFP change is greater than TFP change. The result may have several implications. Does this result relate to the increasing specialisation and/or casualisation of the workforce? Are increasing levels of academic outputs due to more efficient and more effective staff. Could the finding be used to justify increasing labour costs in the system? Or does it reflect capital investment in the system that is helping staff perform duties more efficiently, or in fact replacing some aspects of labour? On the other hand, if staff are becoming more productive because of strategic labour allocation decisions and specialisation, do the results suggest that capital investments are in fact not closely tied to results?

Separate productivity analysis has further cemented the notion that positive productivity change in the sector are being driven by gains in research. It showed that even when inputs can be designated by academic function, research productivity is still shown to grow. At first glance, the finding seems to justify the cross-subsidisation practices raised by participants and in the literature review. There is, of course, no controlling for the value or quality of different types of staff in the separate productivity calculation, however, so most arguments are not yet fully justified. Instead, the finding further highlights the need for accounting systems within universities to specify expenditures by academic function, so that trends can be cross-verified.

The drivers of productivity change that lead to either growth or decline are not consistent across institutions with similar aggregate trends. Change seems to be influenced largely by environmental, strategic and operational contexts. Although many overarching trends are consistent across the system, individual institutional strategies for achieving those trends vary greatly. Efficiency-centric approaches and value creation strategies are apparent in various isolated contexts.

#### 9.2.4b. *Trade-offs with productivity growth*

Research productivity gains appear to have occurred at the expense of education productivity gains in almost all circumstances. The results speak to where innovation—at least innovation with measurable results—in the system is most valued. It may be tempting to discount this finding, citing arguments such as those by Bowen (2012) that the labour-intensive nature of education does not lend itself to standard, technological productivity improvement. This position is discussed in more detail below in Section 9.3.5. It may further be tempting to cite limitations of the measures. The outlier institutions in the dataset, however, call both of these arguments into question. Five institutions' absolute productivity growth was driven by education productivity over the period. Four of them grew research productivity in tandem,

and one appears to have driven education productivity at the expense of research productivity. The prime takeaway is that universities in various circumstances—including one Go8—university significantly increased its productivity by generating more value through its education function. The results call into question the efficacy of the *de facto* system strategy to drive productivity through research.

With respect to the absolute productivity estimates, system trends of research growth appear to be closing gaps between education and research productivity. If the evidence here is taken as valid, then a key question to ask about the sector is, what has been gained and what has been lost from the apparent efforts to converge the relative performance and productivities of education and research? The Go8 institutions are the prime exceptions to these trends. Thus for the non-Go8 intuitions, are they (a) sacrificing past efforts that had led them to excel and specialise in the provision of education services, and has their push to increase research productivity been at the expense of pursuing greater excellence in education? Or, (b) has the push to increase research productivity helped them to become more balanced scholarly institutions that reflect something closer to a desirable academic ideal or a Humboldt style university?

Consider the research productivity gains made by institutions like Australian Catholic University, The University of Notre Dame, and the University of Southern Queensland. Each had education productivity levels estimated as much higher than their research productivity levels at the beginning of the period. What if the same level of strategic commitment had been given to leveraging institutional strengths with even further improvements to practices in learning and teaching? Alternatively, it may be possible that these institutions had reached a threshold for achievement in their provision of education and academic services. Was it the case that to bolster and to achieve more of their mission objectives, they needed to pursue a more thriving teaching-research nexus?

If the estimates of research productivity growth across the system are taken to reflect authentic productivity, then the question of sustainable growth for this function must be entertained. Frequently institutions in the dataset exhibited faster growth in research productivity during the first half of the period. As the as absolute productivity thresholds become narrower, is sustained productivity growth at the current rate possible or desirable? As strong evidence suggests that increases in productivity for one academic function likely come at the expense of increases in the other, how long should institutional leaders and stakeholders desire that such a trend to continue? Worthington and Lee's (2008) study described similar trends in the early 2000s where research productivity growth outpaced that of education. The authors suggested that the improvements in research, rather than teaching, were likely the result of removing inefficiency and were not sustainable. The fact that the trends have sustained, however, raises questions about what has been sacrificed to keep elevating research performance, and portrayals of that performance?

Another question about the sustainability of this trend has to do with whether it is internally or externally driven. Is it the case that increases in research productivity are driven by academic incentives primarily external to individual institutional contexts, as might reflect broader global trends discussed in Chapter 2? If this is the case, then institutional leaders would have to consider their limited control over the phenomenon in trying to either leverage it or temper it. In entertaining these questions, the prime consideration is whether research-

dependent productivity growth reflects true national and stakeholder interests and to what extent the trends are driven by institutional strategies, by global forces or by national policy and incentive frameworks.

#### **9.2.4|d. Australian Education and Research Productivity Paradigms**

The productivity of the Australian higher education system can also be framed conceptually, and the conceptual framings may help facilitate the conversations needed to set desirable strategic directions for the system. In terms of the productivity paradigms discussed above, the operative productivity paradigm for research is ‘value creation productivity’, and the operative paradigm for education is ‘efficiency productivity’. Empirically, the two highest rates of growth for any model component measured in the study are (1) the rate of growth in research outputs and (2) the rate of growth in lower-level, usually fixed-term, teaching-only academic staff. These figures speak volumes in terms of how productivity is approached in Australian universities with respect to education and research. One participant succinctly describes the situation, and others elaborate

*With the situation in Australia, the only place you can grow on the student side is int'l students... So what you need to do is reduce your costs, and do something else with the money. In research, you've got more potential to grow. There's an unlimited market there in some senses.*

–KS.8

*What we've done here is to be quite explicit in separating the research budget from the teaching budget and investing the research budget in the most productive researchers. And then trying to get the most out of the teaching budget. That relates to things like making sure you that you don't have small classes and that you get as much teaching time out of the teaching staff as possible.*

–KS.6

*Most people think of productivity in terms of research output.*

–KS.10

*I'm sure most people think of [productivity] in terms of students first. That is... cost efficient ways of pushing students through the system.*

–KS.7

Finally, to address the issue, another participant offered salient advice.

*Department chairs, deans, and even the provost, need to be engaged in conversations about cost minimisation and mission maximisation. Only through discussion where you see data about your own operations, can you talk about the trade-offs. That's what people have to do down deep in the organisation, and you have to do this over and over again. Every term you have to do it over again. There is no shortcut.*

–TE.9

#### **9.2.4|e. Returns to scale**

Returns to scale in Australian universities is a controversial topic. Australian universities are large and growing. The Australian higher education system is unique in the world with

respect to such a large proportion of its 38 universities following this pattern. Rationales for sustained growth need continuous testing. Are universities growing to achieve economies of scale? What if many institutions are not achieving returns to scale? To what extent is higher education an industry where conventional economies of scale should be expected and targeted?

First true returns to scale in higher education cannot be determined with institution-level indicators. Even though production functions and productivity measurement in the private sector are conventionally used to make determinations about organisational returns to scale, this cannot be done for public universities. As discussed in Chapter 3, the very nature of public institutions and non-for-profit enterprises is that operational surplus gets directed back into the organisation as expenditure on further initiatives to realise mission objectives. Thus, if there are significant returns to scale in certain aspects of the university, they would be hidden in most institution level measures because the margins realised from those activities are systematically redirected back into the organisation either to further enhance the unit that generated the surplus or to cross-subsidise another unit.

Thus, especially with the measured trends of the Go8 institutions, one must take caution in drawing conclusions. At least four separate and unique conclusions could be made. First is the least favourable. Over the period examined, these institutions have grown considerably, yet most of their productivity change trends have marginally decreased. At face value, one could conclude that institutional growth has not led to any scale efficiencies. Alternatively, because their output production is generally very high, and their trends are relatively flat with little room for interpretation, one might conclude that they represent the safest and most productive investment in public funds. Or, one could look at the dynamic trends of most other institutions in the dataset and conclude that, if the government wants to invest in innovation and productivity growth, the Go8 institutions are the worst investment. Then again, another narrative might follow from the paragraph above, in that the Go8 simply finds immediate and strategic ways to reinvest surpluses that they achieve from economies of scale—and they are committed to that cycle—so their scale efficiencies simply are not observable with institution-level metrics.

While there may be considerable truth in the latter narrative, both the qualitative and quantitative findings of this research suggest that there remain opportunities for significant productivity improvement. If the latter narrative serves only to reinforce the status quo and act as an excuse not to reform, then it should not be entertained in policy discussions for improving performance. Rather, the other three interpretations should be given weight, so that more opportunities for intervention can be exposed and discussed.

More broadly with respect to the entire system, evidence remains split on returns to scale. In Australian higher education, at least with respect to the institution-level productivity change trends, there appear to be no general rules or tendencies for the shape of trends with respect to scale. Productivity levels appear to have little to do with the size of institutions, and consistent productivity gains do not appear to be correlated with growth as much as with starting levels of productivity, irrespective of the size of the institution. Evidence suggests that implementing strategies to achieve scale economies in universities is not straight forward. Consider the following participant quotes.

*We're about 70 in the world, and we'd love to be 50. But we have no aim to be 50 because we couldn't afford it. Once you're up there, the cost is outstanding. And then the cost of staying there. You've got your Cal Tech model, where you stay so small, but so elite. But then the investment on an individual faculty member there is outrageous.*

–KS.2

*We find that [universities] are much more efficient the bigger they get... if you look at the actual activities that contribute toward teaching and learning, then you will see economies of scale.*

–TE.2

*Scale economy is a huge issue in higher education... Australian public universities are out to 50,000 students searching for economies of scale... That's the reality. The Australian system is drunk on growth.*

–KS.4

Finally, Participant KS.3 stated, “a key measure of education productivity is the rate at which undergraduates can complete their degree and know that they can ask a senior staff member for a quality recommendation”. If an efficiency productivity paradigm for the education function of universities in Australia remains dominant—that is, as class sizes increase, and as fixed-term, lower level staff deliver most of the lectures—this is precisely the kind of productivity indicator that should expect to diminish.

## **9.2.5 Implications for Policy and Practice**

### **9.2.5|a. Measurement for improvement**

As stressed throughout this dissertation, the purpose of improving higher education productivity measurement is to facilitate the improvement of real productivity and performance. Research findings reinforce knowledge that different stakeholders value aspects of performance differently. Findings also show that measurement tools can be tailored to address varied interests. In addition to the varied portrayals of performance, however, the culmination of observations and trends this research has revealed speaks to broader, national contemporary policy and practice concerns for the system. Findings speak directly to three key policy issues currently under review by the Australian DET, as well as several umbrella issues concerning university operations and best practice. The policy issues are (1) performance funding, (2) provider category standards (PCS), and (3) financial transparency. The issues concerning operations and best practice all relate to structural, operational and ideological constraints to improving productivity.

### **9.2.5|b. Performance funding**

As the current research has shown, seemingly straight forward indicators can be packed with hidden but measurable information that is open to interpretation. Even with robust indicators, different parameterisations, weightings and estimates will yield varying results. Despite the intuitive appeal of performance indicators to inform funding allocations, they may serve only to drive Goodhart's Law: ‘once a measure becomes a target, it ceases to be a good measure’ (Elton, 2004). The prime concerns are that (A) given the malleability of university data,

incentives will serve to boost only reported performance rather than actual performance, or (B) that resources will be diverted from other important areas to meet only the measured performance aspects, sacrificing overall quality. Rationally, however, governments should allocate funds based on available information. Institution-level indicators, though, first need a high resolution of transparency and need to have been vetted as valid and reliable diagnostic instruments.

Muller (2018) describes widespread use of quantitative performance indicators as incentive mechanisms before having been vetted as useful and reliable diagnostic tools. Even if a reliable and agreeable measure of student attrition, for example, is identified as a key performance indicator to guide funding decisions, the indicator itself still says little about why students drop out. Understanding the ‘why’ is what leads to improvement. This research has demonstrated the informative potential of well-considered measures and measurement practice. Importing institution-level indicators into national funding policy guidelines, however, without careful reflection will likely create more funding controversy than it alleviates.

In 2018 the Australian DET announced that future funding growth for public university undergraduate programs would be contingent on meeting performance requirements, with possible criteria listed as student retention and admittance of students from backgrounds of low socio-economic status (DET, 2018c). An open request for proposals outlining appropriate performance funding criteria was issued, and in early 2019 the request for proposals was closed with a model for institutional performance funding set to be finalised in 2019 (DET, 2018d). Rewarding university performance is understandable. Intuitively, when government funds are invested, the public is entitled to ask whether allocations are appropriate and effective. Moore et al. (2019) explains, however, that although governments would love to allocate resources in ways that optimise university operations and performance, institutional management often reflects non-financial arrangement and objectives. Universities’ known tendencies not to be cost-minimising institutions, as discussed in Section 3.3.2, speaks to the complexity of university operations and explains why narrow performance funding structures across the globe have delivered mixed if unremarkable results (Dougherty, Natow, Bork, Jones, & Vega, 2013; Jongbloed, B & Lepori, 2001; Tandberg & Hillman, 2014).

### **9.2.5|c. *Institutional differentiation and diversity***

Institutional differentiation refers to division of function among institutions within a system. There is a longstanding debate around whether institutional differentiation best serves the purposes of higher education (Croxford & Raffe, 2015). Institutional diversity (or variety) may be considered as inherently good and as an indicator of a healthy system. Diversity increases student choice and access (Huisman, Meek, & Wood, 2007). Other stances maintain, however, that differentiation is inextricably linked to stratification (Bastedo & Gumpert, 2003). That is, sanctioned initiatives to promote mission diversity and horizontal differentiation often lead to more hierarchy and vertical differentiation. Concerns are rooted in the idea that if certain functions are valued more than others, then institutional division of function can be undermined. If some institutional functions are considered more high status than others, then the whole system is at risk of mission drift (Croxford & Raffe, 2015). Many contemporary views—as well as evidence from this dissertation—suggest that the widespread

regard for university research has spurred mission drift and institutional homogenisation within the Australian higher education system.

Measurement findings from this research show trends of convergence across the system. They show that gains in research productivity appear to be at the expense of gains in education productivity and that institutional differentiation may be diminishing. The institutional convergence measured from 2007 to 2016 is likely systematic and structural. Structural isomorphism among Australian institutions has been found during earlier periods as well (Croucher & Woelert, 2016). As the heterogeneity of Australian universities looks to be decreasing, the trend appears to be policy-driven. This is the rationale for a current review of Australia's higher education provider category standards (PCS). Coaldrake (2018) explains that Australian PCS are based on the *National Protocols for Higher Education Approval Processes*, which are observed and implemented under the auspices of the Tertiary Education Quality Standards Agency (TEQSA). The PCS delineate the functional differences and requirements between institutions of different types within the system. Institutional diversity, or lack thereof, within Australia has been fuel for recent debate about whether new policy should encourage and support more differentiation within the system.

Coaldrake (2018) notes that Australia stands out among international peers for its codifying of the types and the scale of research required for achieving a legal 'university' designation. The research requirements are argued to form a steep barrier to entry for new potential institutions that could meet diverse students' needs. Further, as independently verified by the APC (2017), incentive structures appear to reward excellence in research more so than excellence in teaching, and such incentives have motivated Australian universities to commonly subsidise research with teaching revenue (APC, 2017). The situation has led critics to challenge the underpinnings of the Australian university, asserting that requirements for new potential institutions to attain 'university' status are too prohibitive.

A view is growing that institutional diversity exists in the Australian system despite the PCS, rather than because of them. Thomson (2019) explains that the PCS do not reflect existing diversity within the system and that, as a regulatory tool, the PCS do not facilitate differentiation.

For governments wishing to facilitate horizontal differentiation, Croxford & Raffe (2015) assert that mechanisms for doing so should start with strengthening aspects of differentiation and diversity that already exist within the system. Based on quantitative and qualitative findings, this appears to be the principal limitation of the PCS. Not only could the Australian PCS be amended to be more conducive to entrance and provision of new HEI forms in Australia, but also they could be tied to current funding arrangements. Tying the PCS to funding arrangements could allow for equitable access to public funds by institutions serving more diverse student needs. The current PCS have no direct relationship to government funding for the provision of teaching or research (Coaldrake, 2018). One participant from this research explained.

*We have not segmented our system.... or the funding requirements. The government hasn't given appropriate parameters... The University of Southern Queensland, for example, should not be regulated in the same way as the University of Melbourne... The US has got some great teaching only universities. Often you'll see their students doing our graduate programs.*

#### 9.2.5)d. *Transparency of data and finances*

Section 2.4.3 of this dissertation discussed problems of trust and power that can arise with mandated release of sensitive data under the pretext of transparency and accountability. In the case of institution-level data on public universities, however, this dissertation supports greater accessibility of information. Improving and maintaining integrated and detailed institutional databases is an issue of both science and practicality.

As the procedures for gathering and cleaning secondary data were not part of the methods of this study, *per se*, the extent of the work to do so was not documented in this dissertation. Efforts to join and integrate the idiosyncratic databases from disparate public sources and prepare data for analysis consumed a substantial amount of time. The efforts required to do so likely form a barrier for researchers on comparable projects with tighter time restrictions. Data that is available publicly could also be more detailed, and there is international precedent for how to do so. The US National Center for Education Statistics maintains a single comprehensive databased containing all comparable data elements used in this study and also disaggregated to a finer level of granularity, especially with respect to finances (NCES, 2019). The NCES IPEDS database also tracks resources allocation to individual academic functions within universities.

The Australian Commonwealth Government is currently working with higher education providers to establish a more transparent framework for reporting institutional financial resource allocation to teaching and research by field of education (DET, 2017). From a scholarly standpoint, more detailed financial accounting systems would allow for more accurate and reliable estimates and portrayals of TFP. From a practical standpoint, there is wide stakeholder interest in better understanding institutional practice for education and research resource allocation. Seven interview participants raised and discussed the practice of Australian universities cross-subsidising research activities with revenue generated from teaching. Participant TE.8 mentioned that “the data exists, it’s just not public.” Another participant found the practice and the lack of transparency around the practice to be misguided. The participant exposed at least one perspective on the downside of employing the efficiency productivity paradigm for education and the value creation paradigm for research.

*The greatest casualty in the equation is the student... the students are casualties in the sense of a system that has robbed resources from education to pay for research.*

–KS.4

Cross-subsidies may, however, be framed in a neutral or a positive light with respect to university operations. Participants TE.5 and KS.2 stated that Australian universities cross-subsidise academic functions to achieve broader mission objectives and to compete on a global scale. Best practice for education and research expenditure, though, does not have to be a secret. More transparency around the marginal costs of education and research activities should not be viewed in a negative light when current controversy surrounding the issue persists precisely because practices are opaque.

### 9.2.5|e. Productivity improvement: ‘Cost Disease’ or ‘Mechanical Turk’?

This section examines rationales both for and against institutional productivity improvement agendas. It highlights the importance of distinguishing between inherent limitations to productivity improvement, reinforcers of the status quo, and technical obstacles—all of which exist as confounding variables within common operational environments. Sullivan et al. (2012) and Bowen (2012) each explain how unbridled productivity agendas in higher education could stoke a ‘race to the bottom’ in terms of quality. Bowen (2012) explains that in most industries, significant productivity improvements are often achieved by substituting capital for labour. In labour-intensive industries, like education however, productivity improvements of this nature are fundamentally limited. The author notes that, while vehicle assembly today can be accomplished with only a fraction of the labour input it required 100 years ago, many key aspects of educational processes remain labour intensive. It is rational to assume that unit costs in education would rise faster than those of other industries. The productivity improvement agenda is thus wedged between competing interests to preserve the fundamentals of formal education and to challenge assumptions that little can be accomplished through innovation.

If only marginal productivity improvements can be expected in education, then leveraging what is possible could be viewed as increasingly important. That is, the small gains that are possible should be understood to the greatest extent possible. The standpoint further supports why it is so important to get measurement right. Using linear programming, Moradi-Motlagh, et al. (2016) claim that Australia’s higher education sector experienced significant productivity gains over the period from 2007 to 2013 at upwards of 15% productivity growth. They explain that the measured gains have been driven by technological progress, and they cite widespread use of IT technologies in both teaching and research as key drivers. Compare this to the results of the current study, which show an approximate 1.5% increase in productivity using the most favourable interpretation of the data. Drivers of technological progress are thus called into question. Productivity measurement and assessment must provide the best quality and most contextualised information possible to help identify true productivity trends and drivers. Productivity assessors should not be content with linear programming techniques designed to generate maximally favourable interpretations of data if those positive results can be used to justify the status quo when real improvement opportunities are hidden behind the methods used to generate estimates.

An anecdote from history provides a compelling metaphor. Standage (2002) documents an 18<sup>th</sup> century hoax, where a clockwork chess playing machine was toured around Europe, fooling professional chess players into believing that the machine could play competitive matches against humans without human intervention. The machine is called the Mechanical Turk. In reality, a human with a view of the chess board controlled the machine from the inside. In modern day, this adage still has relevance. Compare the story to the anecdotes of two participants from this research.

*When I was at [University X], our Vice Chancellor decided that... all [student] applications were going to be taken online. Not a lot of time to change the system, but it was a directive from the top. So, what happened was that on the deadline, applications were taken online, but behind the scenes, there were all these people running around making the system look like it was working... It sounds ridiculous, but that actually happens a lot.*

–KS.9

*IT is an area of inefficiency... Our IT systems aren't well integrated, so we have issues with systems being able to talk easily to one another. There has to be manual intervention in between them.*

–KS.7

Participants framed these issues in terms of typical problems for Australian universities. The quotes above are not, however, intended to suggest that all technological aspects of Australian universities operate difunctionally, nor do they suggest that the problems are inherently long-term. Further, they are not meant to insinuate that only HEIs experience transitional difficulties when rolling out new technological platforms. Yet participant KS.8 laments the lack of student-centric thinking when it comes to back-office processes, explaining that the student does not sympathise with a university's standard operating procedures. Procedures should be changed and made simpler for the benefit of the student, not followed for the sake of satisfying internal convention or protocol. Participant KS.9 elaborated.

*If you talk to anyone that works in the system, they will give you examples of working around the system in order to get their work done. Or working effectively in spite of the systems in place, not because of them.*

–KS.9

Caution must always be taken when forming conclusions about higher education productivity. Yet, interests in the status quo must be distinguished from actual technical parameters. For example, the fact that the SFP estimates in this research are higher than TFP estimates may suggest that staff are working more efficiently and effectively, but money is not being spent as judiciously. Alternatively, the difference could suggest that technological and capital investments are precisely what is enabling people to work more productively. Rationales can always be provided for either pursuing or not pursuing a productivity agenda. There are always winners and losers with the status quo and winners and losers with changes to the status quo. Leaders must balance both longstanding principles and new assessment information when designing and executing productivity and performance improvement initiatives. This includes being more explicit about what is limiting progress in a technical sense, and what is limiting progress in a political sense.

#### **9.2.5|f. Constraints to improvement**

This section explores constraints to improvement. The discussion above has helped to frame broad, contextual institutional productivity issues identified in this research. Recall further that Section 7.3.5 also listed participant-identified top areas for Australian university productivity improvement. Those issues were described as utilisation of facilities, student attrition, academic workload models, and administrative functions. If important and agreeable productivity improvement areas can be identified, then a prime implication for policy and practice is to address the barriers that make productivity improvement difficult to achieve. Table 41 lists the top participant-identified constraints to the above areas for improvement.

**Table 43: Constraints to productivity improvement**

| <b>Hindering factors to Productivity Improvement</b>                | <b>Response Frequency</b> |
|---------------------------------------------------------------------|---------------------------|
| Legacy and mindset                                                  | 9                         |
| Regulatory bodies and structures                                    | 6                         |
| Time horizon for seeing results                                     | 5                         |
| Lack of burning platform and incentives                             | 4                         |
| Neither efficiency nor inefficiency has been systematically defined | 4                         |
| Availability of data                                                | 4                         |

In framing key constraints to improvement, participants of this study were more optimistic than Bowen's (2012) description of higher education's inherent 'cost disease'. Because boundaries were set around areas for productivity improvement, most constraints to improvement were also discussed in terms of challenges that could be overcome. The only unavoidable limitation mentioned is the time horizon between when policy and interventions are made and when definitive output effects could be observed. Not only are sweeping initiatives difficult to implement within institutions with devolved governance structures, but also the longer-term nature of university productive processes means that—from a social science standpoint—linking cause and effect is difficult. By the time output effects are observed, a host of additional intervening and correlated variables are also present.

The only constraint listed in Table 41 that bears further noting is the perception that most Australian universities lack a burning platform or incentives to improve productivity. Legacy, mindset, interests, regulatory structures, and imprecise definitions of productivity have been discussed throughout this dissertation. These issues emerge in the scholarly literature, as well as in everyday institutional decision-making, and a prime argument of this dissertation is to urge scholars and practitioners to incrementally address these constraints. Improper, misaligned or non-existent incentives, however, raise a new issue and reinforce the importance of the previous section on understanding rationales for improving or not improving productivity. One participant suggests that, because of legacy, institutions must have a burning platform to improve productivity, whether that manifests in generating more value from current resources or cutting resource spend to achieve the same output. One participant explains how finances serve as either an incentives or counter-incentives.

*[University X] has a burning platform. Most places do not. [University Y] will never have one... They always bring in a lot of money, and whatever money they get, they spend it. Their margin can appear low, but it doesn't matter. They're very wealthy. So, you can be inefficient in some regards because your income covers it up.*

—KS.1

Another participant provides an alternative and more nuanced perspective, emphasising that surplus allows for experimentation. The participant explains that the majority of institutions with ‘average’ financial situations are the ones that have little incentive to improve productivity. The quote below incidentally mentions the same institutions as the quote above.

*[University Y] has extra funds to be able to try new things and not suffer. [University X] has surplus problems. But [University Z] doesn't have either, so we haven't had any burning platform to change. We haven't had any ultimatum like survive or die... But we also don't have a lot of funds to risk, so there's been a lack of innovation because the pressures aren't there.*

–TE.10

Indeed, the carrots and sticks associated with performance funding policy are intended to artificially impose conditions upon all institutions within the system, which are naturally experienced by universities with financial situations at either extreme. The intuitive appeal of incentivising public university performance is also justified by the non-controversial reality that scarce public funds must be rationed. The paradox of the situation is that, although rationing scarce resources is non-controversial—as opposed to the alternative of keeping all tax revenue locked in a vault or simply not collecting taxes—any specific mechanism for the re-distribution of public funds will be controversial. For every argument that high performance should be rewarded, there is a counter argument that lower performing institutions are the ones in need investment to facilitate change. As long as Australia's universities are supported by public funds and expected to contribute to the public good, there remains strong rationale to support and facilitate the improvement of struggling institutions, rather than letting them fail, as might occur in the private sector with an unsuccessful business. This means that incentive structures cannot be heavy-handed. Rather, they must target contextual and individualised problems. No single funding arrangement or performance indicator will ever be universally accepted. Research into productivity measurement and assessment should, therefore, change scope and direction toward revealing the nuances of performance in context, rather than optimising empirical techniques to examine a field and provide ostensibly definitive, singular portrayals of performance, which are fundamentally unattainable.

## **9.3 Limitations and Future Research**

### **9.3.1 Incorporating Quality**

Productivity results generated in this dissertation were flexible and open to interpretation, especially with respect to the threshold estimation method using ranges of output value weights. These value weights, however, do not capture the full range of higher education quality outputs and processes. And in the absence of true quality representation in these metrics, quality should always be a core part of the productivity conversation.

The extent to which the quality of inputs, processes, and outputs changes over time and varies between institutions is not represented in these results. A baseline for productivity metrics in higher education should include inputs and outputs that are adjusted for quality. As discussed in Chapter 2, however, doing so in a reliable fashion is not currently feasible. In the absence of reliable and transferable quality indicators to be included in productivity metrics, it is crucial to keep tabs on quality in conjunction with productivity to ensure that quality is not being sacrificed for measured productivity gain.

### **9.3.2 Measuring Individual Academic Functions**

A key limitation has been the exclusion in this analysis of the service and engagement function of universities. While most participants expressed that measuring teaching and research outputs was pragmatic and sufficient for a productivity representation, three participants explained that omitting the service and engagement mission component causes problems. The rationale is that a significant portion of institutional inputs may be assumed to support independent service and engagement activities. With respect to Australian universities, participants raised the fact that many traditional academic contracts are written with expectations that 20% of their time and effort should be dedicated to service and engagement.

No consistent or workable solution for the problem emerged from discussions on the topic, however. Nevertheless, at least two study participants referenced objectives from their own institutions for improve the measurement and assessment of service and engagement. Publicly available university annual reports in Australia typically include empirical evidence for progress toward service and engagement objectives. Better accounting of service and engagement output in system-wide fashion, however, would require the establishment of more consistent and reliable indicators. An independent logical framework around what is most commonly reported would need to be developed for Australian universities to inform more inclusive productivity metrics.

A further concern raised by study participants regards relative limitations in measuring education productivity as compared to research productivity. Indicators for the education function have been standardised and have remained largely consistent over time. This includes measures of student program numbers and retention rates. The variety of research metrics, however, is always expanding. There is an argument that because of the expanding suite of potential indicators that one could consider in an assessment—such as ‘likes’, comments, or connections on social media platforms like ResearchGate or Academia and open access publications—the opportunities for portraying growth in research performance are greater than opportunities in education. The current research intentionally considered more traditional and widely used indicators in its portrayals of increasing research productivity, but alternative metrics represent a growing area of interest and enquiry.

### **9.3.3 Interdepartmental Dynamics**

Tracking productivity dynamics at the level of university departments or academic disciplines would likely provide significant explanatory value. Without interdepartmental knowledge, key insights are missed about what is driving the variation between institutions. This may be

especially true for understanding scale economies. To measure cross-subsidisation and to track revenue flows between business and academic units would enrich the analysis. The data required for exposing such dynamics, however, is highly variable and idiosyncratic for individual institutions. University governance structures and operational models evolve in context, and without standardised internal accounting systems, such an effort would require coordinated case studies of individual institutions. To some extent this has been done (Deloitte Access Economics, 2016). Public access to source data, however, is not possible because of issues with ‘commercial in confidence’.

### 9.3.4 Productivity Measurement and Assessment Techniques

Five specific limitations and opportunities are worth noting for future improvement of university productivity measurement techniques, model specifications and assessment methods. Table 44 addresses key issues.

**Table 44: Measurement and assessment limitations**

| Limitation                         | Explanation                                                                                                                                                                                                                                                                                                                            |
|------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Experimental design                | The current empirical research did not involve any hypothesis testing. With the establishment of more reliable and transparent indicators in this research, however, a next step in research could include a more experimental design linking policy to performance.                                                                   |
| SFP input indicators               | In the current study no indicators were included to represent quality or value of individual staff members performing different functions. Any indicator to represent differentials in the value of different types of staff would be contentious. Options though may include salary and benefits, level of hire, or past performance. |
| Student load aggregation           | With available public information, student load measures could be disaggregated by field of education. This could add depth to the analysis. Broad institutional patterns might become clearer with information concerning the scale and success rates for individual fields of education.                                             |
| Student fee revenue                | Just as research output included research income, it could be beneficial to incorporate student revenue data as an education output.                                                                                                                                                                                                   |
| Economic and accounting indicators | Methods for determining university capital services, real rates of return on university capital, and alternatives to the CPI for currency adjustments could improve the precision of results.                                                                                                                                          |

## 9.4 Conclusions and Consequences

### 9.4.1 What has been Learned?

This dissertation showed both the fragility and the value of productivity measurement in higher education. On one hand, the differences in results from successive versions of the model in this research, expose the limitations and uncertainty of aggregate, institution-level

indicators. On the other hand, when the implications of data selection and data treatment within a model are well-understood, alternative productivity portrayals offer an advantage because together they can generate a more cohesive and comprehensive understanding of performance.

The research identified and described different paradigms for understanding and conceptualising productivity. The paradigms guided not only what study participants felt was important to consider in a productivity assessment, but also they shaped what participants identified institutional performance issues, how to overcome them, and what constrains progress. The different paradigms highlight the importance of acknowledging interests and priors when analysing productivity data and when considering it for decision-making.

Findings further included generalisable design principles for developing a higher education productivity measurement model. Current model characteristics were listed that exemplify those principles. Productivity measurement in higher education must be contextually appropriate and fit-for-purpose in the environment where it is intended to be used, but also it must be technically and conceptually sound. The design principles offer an essential set of guidelines for adapting new models to new contexts.

In relation to measurement model characteristics, this dissertation also showed that measurement transparency is an essential characteristic for the development and use of any performance indicator. Learning opportunities from measurement derive from the nature of measurement methods and from the act of conducting the measurement exercise. Learnings derive much less from the precise values of results. Different measurement techniques ostensibly measuring the same phenomenon can lead to drastically different results. The ability to decompose results and understand the drivers of trends is key to a well-considered analysis.

Finally, this dissertation covered implications for three higher education policy areas and two areas of practice. Findings have direct implication for policy discussions on performance-based funding, institutional differentiation, and financial and information transparency. Issues of practice regard the benefits and trade-offs associated with pursuing a productivity improvement agenda within an institution, as well as constraints to improvement within the Australian higher education system.

#### **9.4.2 What are the Consequences?**

Findings from this research elevate the importance of examining and challenging the status quo. Certain key findings may not be surprising. Measured results on Australian higher education align with much public discourse and with the experiential accounts of stakeholders within the system. Notably, findings highlight trade-offs associated with education and research productivity. First, positive productivity portrayals across the sector depend on research-emphasised calculations. Second, improving research productivity appears to be at the expense of improving education productivity. Third, the productivity improvement outlook for research is to create more value, while the productivity improvement outlook for education is to increase efficiency.

The results merit a thought experiment. What if incentive structures were reversed? What if education outputs drew the same kind of quantitative interest and attention as research

outputs? What if companies such as Elsevier and Thomson Reuters put equivalent effort into collecting and aggregating education and student success data? Imagine they started releasing multitudes of real time indicators and comparable metrics on teaching quality per institution and per lecturer? The findings of this dissertation not only reflect a global paradigm regarding perceptions of what is most important, but also they quantify the extent of the trends. Policy makers and stakeholders must decide if the current trajectory is the desired trajectory, or if a different balance is needed for the future of higher education.

The research has further highlighted the need for continued improvement to productivity measurement practices in higher education. One study participant summarises the issue well.

*Efficiency analysis has become a bit of a fad because you can open a software package, you can put data in, and it spits out a beautiful number. And so the push for using DEA is very strong. –TE.2*

This dissertation has been critical of the use of DEA and linear programming techniques for higher education productivity and efficiency analysis. In fact, DEA is a powerful and useful technique, and if used with care, it has the potential to provide unique and valuable insight. For the current state of productivity measurement in higher education, however, it leaves much to be desired. Discussion throughout this dissertation has highlighted its limitations. In a time with increased stakes and greater pressure on funds, research on productivity in higher education needs to shift away from algorithmic methods that prioritise convenience and ease of computation and rather move toward contextualised, fit-for-purpose models.

A closing thought is whether productivity in higher education should be pursued as a means or as an end. Should research and education productivity be treated as a prime institutional objective or as a mechanism for achieving something greater? The implication is not trivial. If better productivity indicators can be developed—that is, indicators with more accuracy and reliability in capturing more important aspects of higher education—then improving productivity might be considered a prime operational objective in achieving broader mission-related aspirations.

Alternatively, as discussed in Chapter 2, it may be best to think of productivity only as a lead indicator, or as a mechanism for positioning institutions to achieve broader, longer-term goals that are not appropriately expressed in annual input-output figures. Answers may vary from institution to institution. As measures improve, though, and as more and higher quality institutional data becomes available, it will be increasingly important to frame the purpose of productivity measurement. More work on this topic is needed than what has been accomplished in these pages, and it is the author's hope that this dissertation can strengthen the foundation of empirical, quantitative work on the topic.

## References

- Abbott, M., & Doucouliagos, C. (2003). The efficiency of Australian universities : a data envelopment analysis. *Economics of Education Review*, 22(1), 89–97.
- Aiello, F., & Ricotta, F. (2016). Firm heterogeneity in productivity across Europe: evidence from multilevel models. *Economics of Innovation and New Technology*, 25(1), 57–89.
- Al-Darrab, I. A. (2000). Relationships between productivity, efficiency, utilization, and quality. *Work Study*, 49(3), 97–104.
- Alach, Z. (2017). The use of performance measurement in universities. *International Journal of Public Sector Management*, 30(2), 102–117.
- Alcaide-López-de-Pablo, D., Dios-Palomares, R., & Prieto, Á. M. (2014). A new multicriteria approach for the analysis of efficiency in the Spanish olive oil sector by modelling decision maker preferences. *European Journal of Operational Research*, 234(1), 241–252.
- Alexander, F. K. (2000). The changing face of accountability. *The Journal of Higher Education*, 71(4), 411–431.
- Altbach, P. G., Reisberg, L., & Rumbley, L. E. (2009). *Trends in Global Higher Education : Tracking an Academic Revolution*. Paris: UNESCO.
- Ammons, D. (2004). Productivity barriers in the public sector. In M. Holzer & L. Seok-Hwan (Eds.), *Public Productivity Handbook* (2nd ed., pp. 139–163). New York, N.Y.: Marcel Dekker, Inc.
- Antle, J. M., & Capalbo, S. M. (1988). An introduction to recent developments in production theory and productivity measurement. In J. M. Antle & S. M. Capalbo (Eds.), *Agricultural productivity: Measurement and explanation* (pp. 17–95). Washington, DC: Johns Hopkins University Press.
- Appelbaum, E. (1978). Testing Neoclassical Production Theory. *Journal of Econometrics*, 7, 87–102.
- Astin, A. W. (2012). *Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education*. Lanham, Md: Rowman & Littlefield Publishers.
- Atkins, J., & Herfel, W. (2006). Counting beans in the degree factory. *International Journal for Educational Integrity*, 2(1).
- Australia Productivity Commission (APC). (2017). *University Education, Shifting the Dial*. Canberra: Commonwealth of Australia.
- Australian Bureau of Statistics (ABS). (2012). Australian National Accounts: Concepts, Sources and Methods, Edition 3. In *Cat. No. 5216.0*.
- Australian Government. *Tertiary Education Quality and Standards Agency Act 2011.* , Pub. L. No. 73, 153 (2011).

- Australian Government. (2017). *The Higher Education Reform Package*. Canberra: The Commonwealth of Australia.
- Australian Qualifications Framework Council. (2013). *Australian Qualifications Framework*. South Australia: The Commonwealth of Australia.
- Australian Research Council. (2015). National Competitive Grants Program. April 6, 2019, Australian Government website: <https://www.arc.gov.au/grants/national-competitive-grants-program>
- Bairam, E. (1994). *Homogeneous and nonhomogeneous production functions: theory and applications*. Avebury.
- Barnett, R. (1992). *Improving higher education : total quality care*. Bristol, PA: Society for Research into Higher Education.
- Baskerville, R., Pries-Heje, J., & Venable, J. (2009). Soft design science methodology. *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology*, 9. ACM.
- Bastedo, M. N., & Gumport, P. J. (2003). Access to what? Mission differentiation and academic stratification in U.S. public higher education. *Higher Education*, 46(3), 341–359.
- Becker, G. S. (1960). Underinvestment in college education? *The American Economic Review*, 2, 346–354.
- Becker, G. S. (1962). Investment in Human Capital: A Theoretical Analysis. *Journal of Political Economy*, 70(6), 9–49.
- Becker, G. S. (1964). *Human capital: a theoretical and empirical analysis, with special reference to education*. New York, N.Y.: Columbia University Press.
- Becker, W. J., & Walstad, R. (2012). *Econometric modeling in economic education research* (2nd ed.). Norwell, MA: Kluwer Academic Publishers.
- Bentley, P. J., & Kyvik, S. (2012). Academic work from a comparative perspective: a survey of faculty working time across 13 countries. *Higher Education*, 63(4), 529–547.
- Besanko, D., Braeutigam, R. R., & Rockett, K. (2015). Inputs and production functions. In *Microeconomics* (5th ed., pp. 183–220). Hoboken, NJ: Wiley.
- Bexley, E., James, R., & Arkoudis, S. (2011). *The Australian Academic Profession in Transition*. Melbourne, Australia: Centre for the Study of Higher Education.
- Bok, D. C. (2005). *Universities in the marketplace: the commercialization of higher education*. Princeton, NJ: Princeton University Press.
- Borden, V. M. H., & Bottrill, K. V. (1994). Performance indicators: History, definitions, and methods. *New Directions for Institutional Research*, 1994(82), 5–17.
- Bornmann, L., & Haunschild, R. (2016). Efficiency of research performance and the glass researcher. *Journal of Informetrics*, 10, 652–654.
- Botsman, R. (2017). *Who Can You Trust?: How Technology Brought Us Together—and Why It Could Drive Us Apart*. Penguin, UK.
- Bowen, W. G. (2012). *The cost disease in higher education: is technology the answer?*

- Stanford, CA: Tanner Lectures, Stanford University.
- Braithwait, S. D. (1980). The substitution bias of the Laspeyres price index: An analysis using estimated cost-of-living indexes. *The American Economic Review*, 70(1), 64–77.
- Brennan, J. (2007). The academic profession and increasing expectations of relevance. In A. Arimoto, J. Brennan, M. J. Finkelstein, J. F. Galaz-Fontes, M. Gil-Antón, L. Padilla-González, ... V. L. Meek (Eds.), *Key challenges to the academic profession* (pp. 19–28). Paris: Kassel, Junior.
- Bureau of Labor Statistics (BLS). (2007). *Technical Information about the BLS Multifactor Productivity Measures*.
- Calderon, A. J. (2018). *Massification of Higher Education Revisited*. Melbourne: RMIT University.
- Carrington, R., Coelli, T. I. M., & Rao, D. S. P. (2005). The performance of australian universities: Conceptual issues and preliminary results. *Economic Papers*, 24(2), 145–163.
- Cave, M., Hanney, S., & Kogan, M. (1991). *The use of performance indicators in higher education : a critical analysis of developing practice*. Jessica Kingsley Publishers.
- Caves, D. W., Christensen, L. R., & Diewert, W. . E. (1982). The Economic Theory of Index Numbers and the Measurement of Input , Output , and Productivity. *Econometrica*, 50(6), 1393–1414.
- Chizmar, J. F., & Zak, T. A. (1983). Modeling Multiple Outputs in Educational Production Function. *Research in Economic Education*, 73(2), 18–22.
- Clark, B. R. (1983). *The higher education system : academic organization in cross-national perspective*. Berkeley, CA: University of California Press.
- Coaldrake, P. (2018). *Review of the Higher Education Provider Category Standards*. Canberra: Department of Education and Training, Australian Government.
- Coates, H. (2016). Reporting alternatives: Future transparency mechanisms for higher education. In E. Hazelkorn (Ed.), *Global rankings and the geo-politics of higher education: Understanding the influence and impact of rankings on higher education, policy and society* (pp. 277–294). London: Taylor and Francis.
- Coates, H. (Ed.). (2017a). *Productivity in Higher Education: Research Insights for Universities and Governments in Asia*. Tokyo: Asia Productivity Organisation.
- Coates, H. (2017b). *The Market for Learning*. Singapore: Springer.
- Coelli, T. (2005). *An Introduction to Efficiency and Productivity Analysis*. New York, N.Y.: Springer.
- Colledge, L., & Verlinde, R. (2014). *SciVal Metrics Guidebook*.
- Cooper, W. W., Seiford, L. M., & Tone, K. (2007). *Data envelopment analysis : a comprehensive text with models, applications, references and DEA-solver software* (2nd ed.). New York, N.Y.: Springer.
- Croucher, G., & Moore, K. (2018). *Eggs and Baskets: Lessons from the 2009 decline in Indian student enrolments*.

- Croucher, G., & Woelert, P. (2016). Institutional isomorphism and the creation of the unified national system of higher education in Australia: an empirical analysis. *Higher Education*, 71(4), 439–453.
- Croxford, L., & Raffe, D. (2015). The iron law of hierarchy? Institutional differentiation in UK higher education. *Studies in Higher Education*, 40(9), 1625–1640.
- De Witte, K., & Sneyers, E. (2017). Efficiency in education: a review of literature and a way forward. *Journal of the Operational Research Society*, 68(4), 339–363.
- Dearden, J. A., Grewal, R., & Lilien, G. L. (2008). *Merit Aid and Competition in the University Marketplace* (Working pa). Department of Economics, College of Business and Economics, Pennsylvania State University.
- Dekeyser, S., Watson, R., & Baré, E. (2016). Comparing academic workload models: How Australian universities resource teaching activities. *Tertiary Education Management Conference: From Rhetoric to Reality*, 1–14.
- Deloitte Access Economics. (2016). *Cost of delivery of higher education*. Canberra: Deloitte.
- Department of Education and Training (DET). (2015). *Higher Education in Australia: A review of reviews from Dawkins to today*. Canberra: Australian Government.
- Department of Education and Training (DET). (2017a). Historical Research Block Grant Programs. April 18, 2019, Australian Government website: <https://www.education.gov.au/historical-research-block-grant-programs>
- Department of Education and Training (DET). (2017b). Increased transparency and accountability. March 11, 2019, The Australian Government website: <https://www.education.gov.au/historical-research-block-grant-programs>
- Department of Education and Training (DET). (2018a). Consolidated Time Series Data. April 6, 2019, Australian Government website: <https://www.education.gov.au/consolidated-time-series-data>
- Department of Education and Training (DET). (2018b). Finance Publication. April 6, 2019, Australian Government website: <https://www.education.gov.au/finance-publication>
- Department of Education and Training (DET). (2018c). Higher Education Policy changes - Provider FAQs. February 24, 2019, Australian Government website: <https://www.education.gov.au/higher-education-policy-changes-provider-faqs>
- Department of Education and Training (DET). (2018d). Performance-Based Funding for the Commonwealth Grant Scheme. February 24, 2019, Australian Government website: <https://www.education.gov.au/performance-based-funding-commonwealth-grant-scheme>
- Department of Education and Training (DET). (2019a). Higher Education Statistics. February 14, 2019, Australian Government website: <http://highereducationstatistics.education.gov.au/>
- Department of Education and Training (DET). (2019b). Research Block Grants Calculation Methodology | Department of Education and Training. May 30, 2019, Australian Government website: <https://www.education.gov.au/research-block-grants-new-arrangements-allocation-calculation-methodology>

- Department of Education and Training (DET). (2019c). Research Income Data (2004-2017). February 14, 2019, Australian Government website: <https://docs.education.gov.au/node/47851>
- Department of Education and Training (DET). (2019d). Research Training Program. April 18, 2019, Australian Government website: <https://www.education.gov.au/research-training-program>
- Disney, R., Haskel, J., & Heden, Y. (2003). Restructuring and Productivity Growth in UK Manufacturing. *The Economic Journal*, 113(489), 666–694.
- Dogramaci, A., & Färe, R. (Eds.). (2012). *Applications of modern production theory: efficiency and productivity*. Dordrecht: Kluwer Academic Publishers Group.
- Dooren, W. van, & Van de Walle, S. (2008). *Performance information in the public sector : how it is used*. New York, N.Y.: Palgrave Macmillan.
- Dougherty, K. J., Natow, R. S., Bork, R. H., Jones, S. M., & Vega, B. E. (2013). Accounting for higher education accountability: Political origins of state performance funding for higher education. *Teachers College Record*, 115(1).
- Drengenberg, N., & Bain, A. (2017). If all you have is a hammer, everything begins to look like a nail – how wicked is the problem of measuring productivity in higher education? *Higher Education Research & Development*, 36(4), 660–673.
- Duan, S. X. (2019). Measuring university efficiency. *Benchmarking: An International Journal*, 26(4), 1161–1173.
- Dulleck, U., & Kerschbamer, R. (2006). On doctors, mechanics, and computer specialists: The economics of credence goods. *Journal of Economic Literature*, 44(March), 5–42.
- Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., & Shale, E. A. (2001). Pitfalls and protocols in DEA. *European Journal of Operational Research*, 132(2), 245–259.
- Ehrenberg, R. G. (2004). Econometric studies of higher education. *Journal of Econometr*, 121, 19–37.
- Elken, M., & Stensaker, B. (2018). Conceptualising ‘quality work’ in higher education. *Quality in Higher Education*, 24(3), 189–202.
- Elsevier B.V. (2018). Benchmarking. SciVal, RELX Group website: <http://www-scival-com.ezp.lib.unimelb.edu.au/benchmarking/analyse>
- Elton, L. (2004). Goodhart’s Law and Performance Indicators in Higher Education. *Evaluation & Research in Education*, 18(1–2), 120–128.
- Fairweather, J. S. (1989). Academic Research and Instruction : The Industrial Connection. *The Journal of Higher Education*, 60(4), 388–407.
- Fallon, D. (1980). *The German university : a heroic ideal in conflict with the modern world*. Colorado Associated University Press.
- Färe, R. (1988). *Fundamentals of production theory*. Berlin: Springer-Verlag.
- Feller, I. (2009). Performance measurement and the governance of American academic science. *Minerva*, 47(3), 323–344.

- Fizel, J., Gustafson, E., & Hadley, L. (1999). *Sports economics: Current research*. Westport, CT: Greenwood Publishing Group.
- Forrester, G. (2011). Performance management in education: milestone or millstone? *Management in Education*, 25(1), 5–9.
- Forrester, J. W. (1993). Low productivity: is it a problem or merely a symptom? In W. F. Christopher & C. G. Thor (Eds.), *Handbook for productivity measurement and improvement*. Cambridge, MA: Productivity Press.
- Fujii, H., Edamura, K., Sumikura, K., Furusawa, Y., Fukuzawa, N., & Managi, S. (2015). How enterprise strategies are related to innovation and productivity change: an empirical study of Japanese manufacturing firms. *Economics of Innovation and New Technology*, 24(3), 248–262.
- Gaynor, A. K. (1998). *Analyzing problems in schools and school systems : a theoretical approach*. Mahwah, N.J.: L. Erlbaum Associates.
- Ghobadian, A., & Husband, T. (1990). Measuring total productivity using production functions. *International Journal of Production Research*, 28(8), 1435–1446.
- Goedegebuure, L. C. J., Maassen, P. A. M., & Westerheijden, D. F. (Eds.). (1990). *Peer review and performance indicators : quality assessment in British and Dutch higher education*. Utrecht: Uitgeverij Lemma.
- Goedegebuure, L., & Marshman, I. (2017). Australian university productivity ; some food for thought. In *Improving Service Sector Productivity: The Economic Imperative* (pp. 69–80). Melbourne: CEDA.
- Gordon, D., & Vaughan, R. (2011). The Historical Role Of The Production Function In Economics And Business. *American Journal of Business Education*, 4(4), 25–30.
- Gralka, S., Wohlrabe, K., & Bornmann, L. (2019). How to measure research efficiency in higher education? Research grants vs. publication output. *Journal of Higher Education Policy and Management*, 1–20.
- Gruber, T. (2014). Academic sell-out: how an obsession with metrics and rankings is damaging academia. *Journal of Marketing for Higher Education*, 24(2), 165–177.
- Gürel, Y. Ü., & Kiliçaslan, Y. (2016). The impact of intangible assets on the productivity of manufacturing firms in Turkey. *METU Studies in Development*, 43(1), 191–227.
- Hansmann, H. (1980). The Role of Non-profit Enterprise. *PONPO Working Paper, Yale University*, 1–67.
- Harland, T., & Wald, N. (2017). Vanilla teaching as a rational choice: the impact of research and compliance on teacher development. *Teaching in Higher Education*, 1–16.
- Hazelkorn, E. (2015). *Rankings and the reshaping of higher education : the battle for world-class excellence*. Basingstoke : Palgrave Macmillan.
- Hevner, A. R. (2007). A Three Cycle View of Design Science Research A Three Cycle View of Design Science Research. *Scandinavian Journal of Information Systems*, 19(2), 1–6.
- Hill, T. (2017). *Manufacturing Strategy: The Strategic Management of the Manufacturing Function*. New York, N.Y.: Palgrave.

- Hoffman, A. J. (1999). Institutional Evolution and Change: Environmentalism and the U.S. Chemical Industry. *Academy of Management Journal*, 42(4), 351–371.
- Holland, D., Liadze, I., Rienzo, C., & Wilkinson, D. (2013). The relationship between graduates and economic growth across countries. *BIS Research Paper*, 110(2013).
- Huisman, J., Meek, L., & Wood, F. (2007). Institutional Diversity in Higher Education: a Cross-National and Longitudinal Analysis. *Higher Education Quarterly*, 61(4), 563–577.
- Izadi, H., Johnes, G., Oskrochi, R., & Crouchley, R. (2002). Stochastic frontier estimation of a CES cost function: the case of higher education in Britain. *Economics of Education Review*, 21, 63–71.
- Jackson, K. (2001). *The Higher Education Contribution Scheme*. Canberra: Parliament of Australia.
- Jaeger, D. A., & Page, M. E. (1996). Degrees Matter: New Evidence on Sheepskin Effects in the Returns to Education. *The Review of Economics and Statistics*, 78(4), 733–740.
- James, E., & Neuberger, E. (1981). The university department as a non-profit labor cooperative. In *Collective choice in education* (pp. 237–234). Dordrecht: Springer.
- Johnes, G., & Johnes, J. (2009). Higher education institutions' costs and efficiency: Taking the decomposition a further step. *Economics of Education Review*, 28(1), 107–113.
- Johnes, J. (2006). Data envelopment analysis and its application to the measurement of efficiency in higher education. *Economics of Education Review*, 25(3), 273–288.
- Johnson, J. L. (1978). The role of the student in the higher education production function. *Research in Higher Education*, 9(2), 169–179.
- Jongbloed, B., & Lepori, B. (2001). Keeping up performances: An international survey of performance-based funding in higher education. *Journal of Higher Education Policy and Management*, 23(2), 127–145.
- Kallio, K., & Kallio, T. (2014). Management-by-results and performance measurement in universities - implications for work motivation. *Studies in Higher Education*, 39(4), 574–589.
- Knight, J. (2003). GATS, trade and higher education: perspective 2003—where are we? In *The Observatory on Borderless Higher Education*, London.
- Kogan, M., & Teichler, U. (2007). *Key challenges to the academic profession*. Mönchebergstraße: Kassel, Jenior.
- Kokko, A. (1996). Productivity spillovers from competition between local firms and foreign affiliates. *Journal of International Development*, 8(4), 517–530.
- Kumbhakar, S., & Lovell, C. A. K. (2003). *Stochastic frontier analysis*. Cambridge University Press.
- Lee, B. L., & Worthington, A. C. (2016). A network DEA quantity and quality-orientated production model: An application to Australian university research services. *Omega*, 60, 26–33.
- Letizia, A. (2015). Revitalizing Higher Education and the Commitment to the Public Good: A

- Literature Review. *InterActions: UCLA Journal of Education and Information Studies*, 11(2), 1–20.
- Liefner, I. (2003). Funding, resource allocation, and performance in higher education systems. *Higher Education*, 46(4), 469–489.
- Lin, B., & Xie, C. (2014). Energy substitution effect on transport industry of China-based on trans-log production function. *Energy*, 67, 213–222.
- Lin, J. (2012). A case study in anti-gaming mechanisms for altmetrics: PLOS ALMs and DataTrust. *Altmetrics 12 ACM Web Science Conference*.
- Linna, P., Pekkola, S., Ukko, J., & Melkas, H. (2010). Defining and measuring productivity in the public sector: managerial perceptions. *International Journal of Public Sector Management*, 23(3), 300–320.
- Marginson, S. (1993). *Education and public policy in Australia*. Cambridge, England: Cambridge University Press.
- Marginson, S. (2013). The Impossibility of Capitalist Markets in Higher Education. *Journal of Education Policy*, 28(3), 353–370.
- Marginson, S., & van der Wende, M. (2007). *Globalisation and Higher Education*. Paris: OECD Education Working Papers, No. 8. OECD Publishing.
- Massy, W. F. (2003). *Honoring the trust: quality and cost containment in higher education*. Williston, VT: Publishing Company.
- Massy, W. F. (2012). Productivity Indices for Nonprofit Enterprises. *Working Paper*, 1-13. Florence, MA.
- Massy, W. F. (2016). *Reengineering the University: How to Be Mission Centered, Market Smart, and Margin Conscious*. Baltimore: The Johns Hopkins University Press.
- Massy, W. F. (2017). *Higher education productivity metrics: an essential guide for users and developers*. Tokyo: Asia Productivity Organisation.
- Massy, W. F., & Archer, S. (2018). Perspectives and advances on productivity measurement in higher education. In E. Hazelkorn, H. Coates, & A. McChormick (Eds.), *Research handbook on quality, performance and accountability in higher education* (pp. 70–80). Edward Elgar Publishing.
- Mawson, P., Carlaw, K. I., & Mclellan, N. (2003). *Productivity measurement: Alternative approaches and estimates*. Wellington, NZ: New Zealand Treasury.
- McBurney, D., & White, T. L. (2010). *Research methods* (8th ed.). Wadsworth Cengage Learning.
- Mihaljevic, M. (2008). Examining the Efficiency of Croatian Higher Education: An Application of Student Attainment Modeling. *Eastern European Economics*, 46(1), 74–88.
- Miles, P. C., Peterson, M., Miles, G., & Bement, D. (2018). Higher education: exploring productivity over time education. *Journal of Applied Research in Higher Education*, 10(4), 534–546.
- Miller, B. A. (2007). *Assessing organizational performance in higher education*. San

- Fransisco, CA: John-Wiley & Sons.
- Mishra, S. K. (2010). A Brief History of Production Functions. *IUP Journal of Managerial Economics*, 8(4), 6–34.
- Misterek, S. D. A., Dooley, K. J., & Anderson, J. C. (1992). Productivity as a Performance Measure. *International Journal of Operations & Production Management*, 12(1), 29–45.
- Monk, D. H. (1989). The Education Production Function: Its Evolving Role in Policy Analysis. *Educational Evaluation and Policy Analysis*, 11(1), 31–45.
- Moore, K. (2018). Improving Higher Education Productivity and Its Measurement: Linking Productivity and Student Success in Australia. *International Journal of Chinese Education*, 7(1), 107–128.
- Moore, K., Coates, H., & Croucher, G. (2018a). Investigating applications of university productivity measurement models using Australian data. *Studies in Higher Education*, 1–15.
- Moore, K., Coates, H., & Croucher, G. (2018b). Understanding and improving higher education productivity. In E. Hazelkorn, H. Coates, & A. C. McCormick (Eds.), *Research Handbook on Quality, Performance and Accountability in Higher Education*. In Press: Edward Elgar Publishing.
- Moore, K., Coates, H., & Croucher, G. (2019). Measuring international higher education productivity: lessons from nine countries in Asia. *Higher Education Forum*, 16(March), 69–84.
- Moore, K., Croucher, G., & Coates, H. (2019). Productivity and Policy in Higher Education. *Australian Economic Review*, 52(2), 236–247.
- Moradi-Motlagh, A., Jubb, C., & Houghton, K. (2016). Productivity analysis of Australian universities. *Pacific Accounting Review*, 28(4), 386–400.
- Muller, J. Z. (2018). *The tyranny of metrics*. Princeton, NJ: Princeton University Press.
- National Center for Education Statistics (NCES). (2019). Integrated Postsecondary Education Data system. Institute for Education Sciences, Washington D.C. website: <https://nces.ed.gov/ipeds/>
- National Institute of Standards and Technology (NIST). (2008). Metrics and Measures. May 30, 2019, US Commerce Department website: [https://samate.nist.gov/index.php/Metrics\\_and\\_Measures.html](https://samate.nist.gov/index.php/Metrics_and_Measures.html)
- Neumann, R. (1996). Researching the Teaching-Research Nexus: A Critical Review. *Australian Journal of Education*, 40(1), 5–18.
- Nicholls, G. (2001). *Professional development in higher education : new dimensions and new directions*. London: Routledge.
- Norton, A., & Cherastidtham, I. (2018). *Mapping Australian higher education 2018*. Melbourne: Grattan Institute.
- Olley, G. S., & Pakes, A. (1992). *The Dynamics of Productivity in the Telecommunications Equipment Industry*. Cambridge, MA: National Bureau of Economic Research.
- Olssen, M., & Peters, M. A. (2005). Neoliberalism, Higher Education and the Knowledge

- Economy: From the Free Market to Knowledge Capitalism. *Journal of Education Policy*, 20(3), 313–345.
- Organisation for Economic Cooperation and Development (OECD). (2001). *Measuring Productivity: Measurement of Aggregate and Industry-level Productivity Growth*. Paris.
- Paige, R. M. (2005). Internationalization of higher education: Performance assessment and indicators. *Nagoya Journal of Higher Education*, 5(8), 99–122.
- Park, J. H. (1999). Estimation of sheepskin effects using the old and the new measures of educational attainment in the Current Population Survey. *Economics Letters*, 62(2), 237–240.
- Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: an integrated approach*. Lawrence Erlbaum Associates.
- Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3), 45–77.
- Plattner, H., Meinel, C., & Leifer, L. (Eds.). (2016). *Design Thinking Research*. Cham: Springer International Publishing.
- Plomp, T., & Nienke, N. (Eds.). (2007). *An Introduction to Educational Design Research*. Enschede: SLO Netherlands Institute for Curriculum Development.
- Porter, M. E. (2008). *Competitive advantage: Creating and sustaining superior performance*. New York, N.Y.: The Free Press.
- Pritchard, R. D. (1995). *Productivity measurement and improvement: Organizational case studies*. Westport, CT: Greenwood Publishing Group.
- Psacharopoulos, G. (1970). Estimating shadow rates of return to investment in education. *The Journal of Human Resources*, 5(1), 34–50.
- Psacharopoulos, G. (1972). Rates of Return to Investment in Education around the World. *Comparative Education Review*, 16(1), 54–67.
- Ray, S. C. (2004). *Data envelopment analysis: theory and techniques for economics and operations research*. Cambridge University Press.
- Reeves, T. (2006). Design research from a technology perspective. In J. van den Akker, K. Gravemeijer, S. McKenny, & N. Nieveen (Eds.), *Educational design Research*. London: Routledge.
- Rowe, P. G. (1991). *Design Thinkng*. Cambridge, England: The MIT Press.
- Said, A., & Torra, V. (Eds.). (2019). *Data Science in Practice*. Cham: Springer International Publishing.
- Santiago, P., Tremblay, K., Basri, E., & Arnal, E. (2008). *Tertiary education for the knowledge society* (Volume 1 a). Paris: OECD Publishing.
- Scerri, M., & Agarwal, R. (2014). Service enterprise productivity in action (SEPIA). In A. Emrouznejad & E. Cabanda (Eds.), *Managing Service Productivity: Using Frontier Efficiency Methodologies and Multicriteria Decision Making for Improving Service Performance* (Vol. 215, pp. 93–114). Heidelberg New York Dordrecht London:

Springer.

- Schiffbauer, M., & Ospina, S. (2010). *Competition and firm productivity: Evidence from firm-level data* (No. 10-67). International Monetary Fund.
- Schofer, E., & Meyer, J. W. (2005). Education in the Twentieth Century. *American Sociological Review*, 70(December), 898–920.
- Schultz, T. W. (1960). Capital Formation by Education. *Journal of Political Economy*, 68(6), 571–583.
- Schultz, T. W. (1961). Investment in Human Capital. *The American Economic Review*, 51(1), 1–17.
- Scott, P. (2006). The academic profession in a knowledge society. *Wenner Gren International Seiries*, (83), 19–30.
- Scott, W. R. (2008). *Institutions and organizations : ideas and interests*. Thousand Oaks, CA: Sage Publications.
- Shavelson, R. J., Phillips, D. C., Towne, L., & Feuer, M. J. (2003). On the Science of Education Design Studies. *Educational Researcher*, 32(1), 25–28.
- Singh, H., Motwani, J., & Kumar, A. (2000). A review and analysis of the state-of-the-art research on productivity measurement. *Industrial Management & Data Systems*, 100(5), 234–241.
- Sink, D. S., & Tuttle, T. C. (1989). *Planning and measurement in your organization of the future*. Industrial Engineering and Management Press.
- Slack, N., Chambers, S., & Johnston, R. (2010). *Operations Management* (6th ed.). Harrow, England: Prentice Hall.
- Spence, M. (1973). Job Market Signaling. *The Quarterly Journal of Economics*, 87(3), 355–374.
- Standage, T. (2002). *The mechanical Turk: the true story of the the chess-playing machine that fooled the world*. Allen Lane: The Penguin Press.
- Stufflebeam, D. L. (2003). The CIPP Model for Evaluation. In *International Handbook of Educational Evaluation* (pp. 31–62). Dordrecht: Springer Netherlands.
- Sullivan, T. A., Mackie, C., Massy, W. F., & Sinha, E. (2012). *Improving Measurement of Productivity in Higher Education*. Washington, DC: National Academies Press.
- Sumanth, D. J. (1994). *Productivity engineering and management*. McGraw-Hill.
- Szuwarzyński, A. (2019). Benefit of the doubt approach to assessing the research performance of Australian universities. *Higher Education Quarterly*, 73(2), 235–250.
- Tandberg, D. A., & Hillman, N. W. (2014). State higher education performance funding: Data, outcomes, and policy implications. *Journal of Education Finance*, 39(3), 222–243.
- Tangen, S. (2005). Demystifying productivity and performance. *International Journal of Productivity and Performance Management*, 54(1), 34–46.
- Teichler, U., Arimoto, A., & Cummings, W. K. (2013). *The Changing Academic Profession: major findings of a comparative survey*.

- Tellis, G. J., & Johnson, J. (2007). The Value of Quality. *Marketing Science*, 26(6), 758–773.
- Thomson, V. (2019). *Go8 Submission: Provider Category Standards*. Canberra: Group of Eight Australia.
- Titus, M. A., & Eagan, K. (2016). Examining Production Efficiency in Higher Education: The utility of stochastic frontier analysis. In M. Paulsen (Ed.), *Higher Education: Handbook of Theory and Research* (Vol. 31, pp. 441–512).
- Tran, C. T. T., & Crawford, M. (2017). Efficiency Based Ranking: A Case of Vietnamese Public Universities. *Asian-African Journal of Economics and Econometrics*, 17(2), 197–215.
- Trow, M. (1973). *Problems in the Transition from Elite to Mass Higher Education*. Berkeley, CA: Carnegie Commission on Higher Education.
- Trow, M. (1999). From Mass Higher Education to Universal Access: The American Advantage. *Minerva: A Review of Science, Learning & Policy*, 37(4), 303–328.
- Trow, M. (2007). Reflections on the Transition from Elite to Mass to Universal Access: Forms and Phases of Higher Education in Modern Societies since WWII. In *International Handbook of Higher Education* (pp. 243–280).
- Tsionas, E. G., Malikov, E., & Kumbhakar, S. C. (2018). An internally consistent approach to the estimation of market power and cost efficiency with an application to U . S . banking. *European Journal of Operational Research*, 270, 747–760.
- UNESCO. (1998). *World Declaration on Higher Education for the Twenty-first Century: Vision and Action*. Paris: World Conference on Higher Education.
- UNESCO. (2017). *Six ways to ensure higher education leaves no one behind; Global education monitoring report* (Vol. 30). UNESCO.
- Van Damme, D. (2015). Global higher education in need of more and better learning metrics. Why OECD's AHELO project might help to fill the gap. *European Journal of Higher Education*, 5(4), 425–436.
- van Vught, F., & Ziegele, F. (Eds.). (2012). *Multidimensional ranking: the design and development of U-Multirank*.
- Vingaard Johansen, U., Knudsen, F. B., Engelbrecht Kristoffersen, C., Stellfeld Rasmussen, J., Saaby Steffen, E., & Sund, K. J. (2017). Political discourse on higher education in Denmark: from enlightened citizen to homo economicus. *Studies in Higher Education*, 42(2), 264–277.
- Watson, R., King, R., Dekeyser, S., Bare, L., & Baldock, C. (2015). Current practice in academic workload allocation processes in Australia. *Tertiary Education and Management Conference 2015: Leading Locally Competing Globally (TEMC 2015)*. Wollongong.
- Welch, F. (1970). Education in Production. *Journal of Political Economy*, 78(1), 35–59.
- Williams, J. (2016). A critical exploration of changing definitions of public good in relation to higher education. *Studies In Higher Education*, 41(4), 619–630.
- Williamson, K., & Johanson, G. (Eds.). (2013). *Research Methods: Information, Systems and Contexts* (1st Editio). Prarhan, Australia: Tilde University Press.

- Woelert, P., & Yates, L. (2015). Too little and too much trust: performance measurement in Australian higher education. *Critical Studies in Education*, 56(2), 175–189.
- Wolszczak-derlacz, A., & Parteka, J. (2013). Dynamics of productivity in higher education : cross-european evidence based on bootstrapped Malmquist indices. *Journal of Productivity Analysis*, 40(1), 67–82.
- Wolszczak-derlacz, J. (2017). An evaluation and explanation of (in) efficiency in higher education institutions in Europe and the US with the application of two-stage semi-parametric DEA. *Research Policy*, 46(9), 1595–1605.
- Yaohua, R., Muyu, L., Weihu, C., & Xianyu, C. (2018). Efficiency, Technology and Productivity Change of Higher Educational Institutions Directly under the Ministry of Education of China in 2007-2012. *Procedia Computer Science*, 139, 598–604.
- Yousif, A. S. H., & Dale, B. . (1990). The influence of inflation on productivity calculations: a case study. *Engineering Costs and Production Economics*, 20, 13–21.

## Appendix A

### Model usage guidelines for APO Practitioners

**Table 45: Input-output framework for APO model test**

| Indicator         | Variable                                    | Data description                                      |
|-------------------|---------------------------------------------|-------------------------------------------------------|
| Financial inputs  | Labor                                       | Academic staff salary and benefits ( $L_1$ )          |
|                   |                                             | Non-academic staff salary and benefits ( $L_2$ )      |
|                   | Capital                                     | Land capital services ( $K_1$ )                       |
|                   |                                             | Buildings capital services ( $K_2$ )                  |
|                   |                                             | Equipment and other capital services ( $K_3$ )        |
|                   |                                             | Repairs and maintenance ( $K_4$ )                     |
|                   | Intermediaries                              | Grants and scholarships ( $I_1$ )                     |
| Education outputs | Administration and other expenses ( $I_2$ ) |                                                       |
|                   | Student load                                | Number of full-time coursework students ( $E_1$ )     |
|                   | Coursework completions                      | Number of coursework graduates ( $E_2$ )              |
|                   | Graduate employment                         | Proportion of prior year graduates employed ( $E_3$ ) |
| Research outputs  | Learning Outcomes                           | Proportion of learning outcomes achieved ( $E_4$ )    |
|                   | Publications                                | Number of publications ( $R_1$ )                      |
|                   | Citations                                   | Number of new citations ( $R_2$ )                     |
|                   | Patents                                     | Number of patents ( $R_3$ )                           |
|                   | Research completions                        | Number of research graduates ( $R_4$ )                |
|                   | Research funds                              | Amount of research funding ( $R_5$ )                  |

Practitioners from participating countries were initially advised to use a calculation of adjusted load, following the work advanced by the United States NRC (Sullivan, et al., 2012). The additional education constructs provided for in the expanded output framework, however, meant that the calculation for the education index  $E_{\Delta t}$  needed revision from that of the NRC model. The practitioners were provided with a tool in Microsoft Excel to make the calculations automatically, as long as they provided the data. Under the assumption that all indicators are used in the calculation, the education component index  $E_{\Delta t}$  from time  $t - 1$  to time  $t$  is given by the following.

$$E_{\Delta t} = \left( \frac{E_{1,t} + E_{2,t}}{E_{1,t-1} + E_{2,t-1}} \right)^{v_{E1,2}} * \left( \frac{E_{3,t}}{E_{3,t-1}} \right)^{v_{E3}} * \left( \frac{E_{4,t}}{E_{4,t-1}} \right)^{v_{E4}}, \quad [15]$$

where the baseline assumption is  $v_{E_i} = \frac{1}{k-1} = \frac{1}{3}$ . And generally,  $\sum_i^{k-1} v_{E_i} = 1$ , where  $k$  is the number of education output components used in model.

Under the assumption that all indicators are used in the calculation, the research component index  $R_{\Delta t}$  from time  $t - 1$  to time  $t$  is given by the following.

$$R_{\Delta t} = \left( \frac{R_{1,t}}{R_{1,t-1}} \right)^{v_{R1}} * \left( \frac{R_{2,t}}{R_{2,t-1}} \right)^{v_{R2}} * \left( \frac{R_{3,t}}{R_{3,t-1}} \right)^{v_{R3}} * \left( \frac{R_{4,t}}{R_{4,t-1}} \right)^{v_{R4}} * \left( \frac{R_{5,t}}{R_{5,t-1}} \right)^{v_{R5}}, \quad [16]$$

where  $v_{R_i} = \frac{1}{m} = \frac{1}{5}$ . And generally,  $\sum_i^m v_{R_i} = 1$ , where  $m$  is the number of research output components used in the model.

## Appendix B

### *Cumulative Productivity Change Calculation Details*

Let  $P_{t_0,1}$  represent the calculation of productivity change from time  $t_0$  to time  $t_1$ . Recall from the previous chapter that  $P_{t_0,1} = \frac{Y_{t_0,1}}{X_{t_0,1}}$ . For simplicity, consider the simplest Törnqvist productivity change index with one output element and one input element. The indexes  $Y_{\Delta t}$  and  $X_{\Delta t}$  normally signify the aggregation of multiple outputs and inputs, but here we demonstrate the simplest case without loss of generality. The following equation illustrates the case of one output and one input.

$$P_{t_0,1} = \frac{Y_{t_0,1}}{X_{t_0,1}} = \frac{\frac{Y_{t_1}}{Y_{t_0}}}{\frac{X_{t_1}}{X_{t_0}}} \quad [17]$$

Now consider a period with multiple time points from  $t_0$  to  $t_n$ . The cumulative productivity change index  $\Pi P_{\Delta t}$  from  $t_0$  to  $t_n$  is given by the iterative product  $\prod_0^{n-1} P_{t_j,j+1}$ . The result of the iterative product is equivalent to calculating a single productivity change index  $P_{\Delta t}$  using data from only the first time point and the final time point in the period, i.e. performing the calculation  $P_{t_0,n}$ . The following equation illustrates the identity.

$$\prod_0^{n-1} P_{t_j,j+1} = P_{t_0,1} \times P_{t_1,2} \times \dots \times P_{t_{n-1},n} = \frac{\frac{Y_{t_1}}{Y_{t_0}} \times \frac{Y_{t_2}}{Y_{t_1}} \times \dots \times \frac{Y_{t_n}}{Y_{t_{n-1}}}}{\frac{X_{t_1}}{X_{t_0}} \times \frac{X_{t_2}}{X_{t_1}} \times \dots \times \frac{X_{t_n}}{X_{t_{n-1}}}} = \frac{\frac{Y_{t_n}}{Y_{t_0}}}{\frac{X_{t_n}}{X_{t_0}}} = P_{t_0,n} \quad [18]$$

And thus, the vector of  $\Pi P_{\Delta t}$  values for institution  $i$  over the period  $t_0$  to  $t_n$  is given by:

$$\Pi P_{t_0,n}^i = [1, P_{t_0,1}^i, P_{t_0,2}^i, \dots, P_{t_0,n}^i] \quad [19]$$

## Appendix C

### *Key Stakeholder Pre-Interview Info Packet*

#### **Australian higher education productivity**

With the amount of available data on the Australian higher education system, metrics and proxies for the sector's productivity are not difficult to generate. Two important questions, however, remain unanswered: (1) to what extent are current metrics able to capture the most important aspects of higher education productivity, and (2) how should metrics be improved to better inform high-stakes decision-making?

#### **Productivity measurement**

Productivity across industries is widely understood to be the ratio of a firm's outputs produced to its inputs consumed,  $\frac{O}{I}$ . Thus, an improvement in this ratio signifies an improvement or innovation in the firm's production process, as long as output quality does not drop. This same logic can be applied to higher education institutions and systems. The following model shows how currently available Australian higher education indicators can be organized into institutional inputs and outputs.

#### **Sample productivity model for Australian higher education**

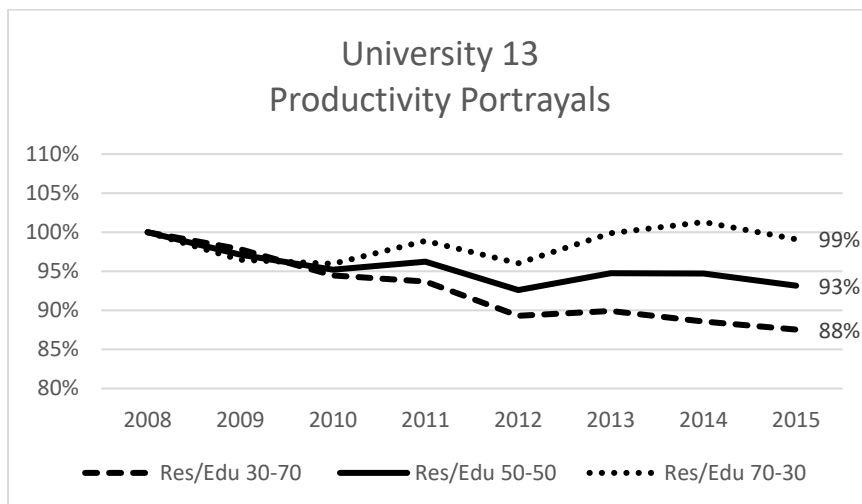
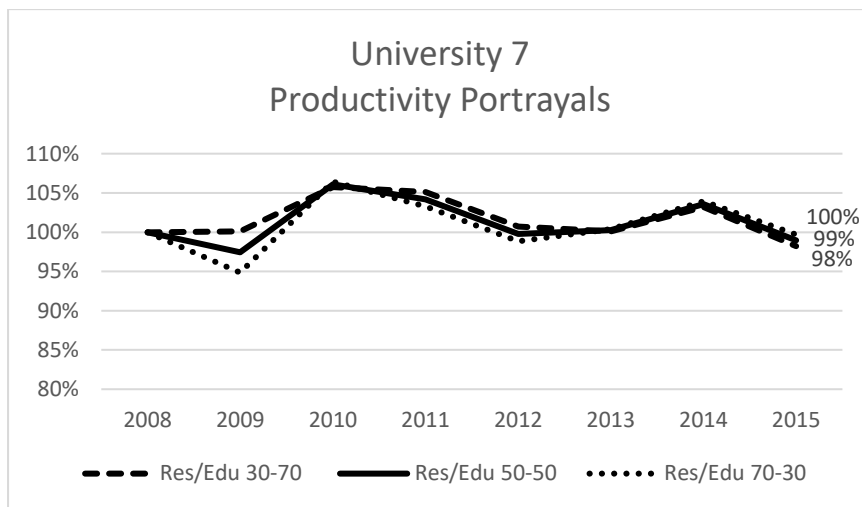
| <b>Inputs</b>  | <b>Data elements</b>                   |
|----------------|----------------------------------------|
| Labor          | Academic staff salary and benefits     |
|                | Non-academic staff salary and benefits |
| Capital        | Land                                   |
|                | Buildings                              |
|                | Equipment                              |
| Intermediaries | Grants and scholarships                |
|                | Administration and other expenses      |

| <b>Education Outputs</b> | <b>Data elements</b>                    | <b>Output 'Value' Weights</b> |
|--------------------------|-----------------------------------------|-------------------------------|
| Student load             | Number of full-time coursework students | 50%                           |
| Coursework completions   | Number of coursework graduates          |                               |
| <b>Research Outputs</b>  |                                         |                               |
| Publications             | Number of publications                  | 50%                           |
| Research completions     | Number of research graduates            |                               |
| Research funds generated | Amount of research funding              |                               |

No single representation of higher education productivity may satisfy every stakeholder; however, there exist better portrayals and worse portrayals of an institution's productivity over time. Both internal operational factors and stakeholder values should be taken into consideration when generating a characterisation of an institution's productivity. Such can be

achieved in a more representative fashion if different interpretations of the data are considered.

## Different representations of Australian university productivity data



The figures above illustrate how productivity trends in higher education can appear different, depending on (A) institutional operational contexts and (B) the perceived impact of their activities. The trends in each figure vary because of different interpretations of how much ‘value’ research and teaching activities create, respectively. A key finding of the exercise highlights the need for further discussion about how the value of—especially—non-monetary outputs of higher education should be portrayed in a productivity analysis for individual institutions and across a system.

## Appendix D

### *Technical Expert Pre-Interview Info Packet*

#### **Australian higher education productivity**

With the amount of available data on the Australian higher education system, metrics and proxies for the sector's productivity are not difficult to generate. Two important questions, however, remain: (1) to what extent are current metrics able to capture the most important aspects of higher education productivity, and (2) how should metrics be improved to better inform high-stakes decision-making?

#### **Productivity measurement**

Productivity across industries is widely understood to be the ratio of a firm's outputs produced to its inputs consumed,  $\frac{O}{I}$ . Thus, an improvement in this ratio signifies an improvement or an innovation in the firm's production process—as long as output quality does not drop. This same logic can be applied to higher education institutions and systems. The following model shows how currently available Australian higher education indicators can be organized into institutional inputs and outputs.

#### **Sample productivity model for Australian higher education**

| Inputs         | Data elements                          |
|----------------|----------------------------------------|
| Labor          | Academic staff salary and benefits     |
|                | Non-academic staff salary and benefits |
| Capital        | Land ( <i>rental value</i> )           |
|                | Buildings ( <i>rental value</i> )      |
|                | Equipment ( <i>rental value</i> )      |
| Intermediaries | Grants and scholarships                |
|                | Administration and other expenses      |

| Education Outputs        | Data elements                           | Output 'Value' Weights |
|--------------------------|-----------------------------------------|------------------------|
| Student load             | Number of full-time coursework students | 50%                    |
| Coursework completions   | Number of coursework graduates          |                        |
| <b>Research Outputs</b>  |                                         |                        |
| Publications             | Number of publications                  | 50%                    |
| Research completions     | Number of research graduates            |                        |
| Research funds generated | Amount of research funding              |                        |

No single representation of higher education productivity will satisfy every stakeholder; however, there exist better portrayals and worse portrayals of an institution's productivity over time. Both internal operational factors and stakeholder values should be taken into consideration when generating a characterisation of an institution's productivity.

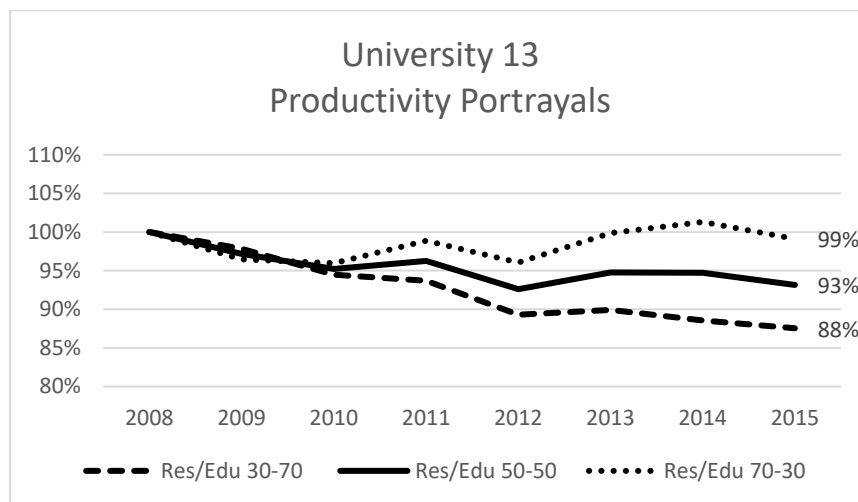
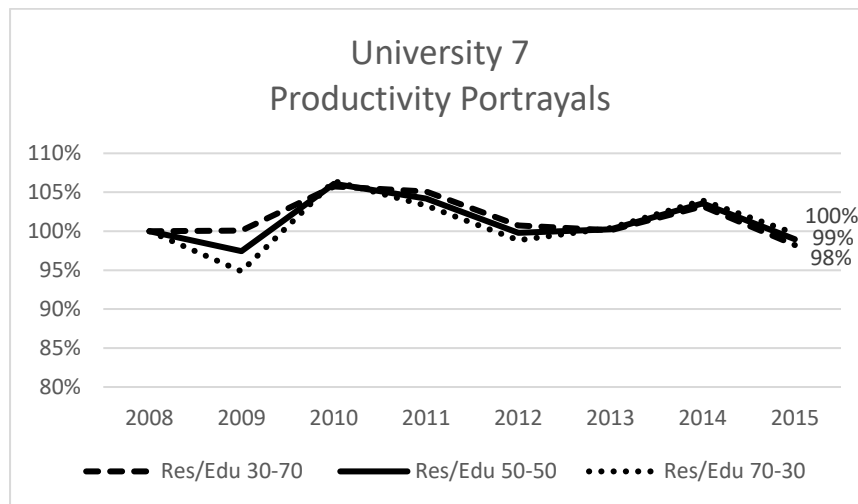
#### **Input & output aggregation**

The current research uses Törnqvist indexes for aggregating both inputs and outputs to calculate productivity metrics. A Törnqvist higher education output index is given by:

$$Q_t = \prod_{i=1}^n \left( \frac{q_{i,t}}{q_{i,t-1}} \right)^{u_i},$$

where  $n$  is the number of outputs an institution produces;  $q_{i,t}$  is the amount of output  $i$  at time  $t$ , and  $q_{i,t-1}$  is the amount of output  $i$  at time  $t - 1$ . The variable  $u_i < 1$  represents an *estimated* value weight of output  $i$ . It may be interpreted as the proportion of value created by output  $i$  with respect to the total value created by all outputs. Thus,  $\sum_i^n u_i = 1$ . The default assumption of the current model is that education outputs and research outputs are weighted equally, 50-50. For Törnqvist input indexes, the process of determining  $u_i$  is more systematic. The value of  $u_i$  for input  $i$  is the proportion of annual expenditure on input  $i$ , with respect to the total expenditure on all inputs.

### Different representations of Australian university productivity over time



The figures above illustrate how productivity trends for Australian universities may appear different depending on (A) interpretations of the value created by their outputs, and (B) institutional context. The discrepancies between trends for a single institution are the result of assigning different value weights to the education and research outputs. Findings highlight

the need for further discussion about how both monetary and non-monetary outputs should be represented in productivity calculations.

## Appendix E

### *Interview Discussion Guide*

**Participant Info:** Please state your name and professional roles.

*##Script: My explanation of research goals, background, motives and objectives*

**Objective 1:** Advance a more precise understanding of what is meant by higher education productivity.

**Objective 2:** Develop an appropriate and reliable model for indicating higher education productivity that can inform high-stakes decision-making. Develop a model that “measures what is valuable—not one that values what is measurable”.

**Question 1:** How do you view productivity in higher education and what does it mean to you?

*## Script: If I could draw your attention to the materials I’ve provided. This is my working model and characterization of higher education productivity. Perhaps I can briefly explain what the model shows.*

**Question 2:** How would you react to the characterization of productivity that I have provided you in the model? Specifically, I’m interested in what you think about the ‘**inputs**’ and ‘**outputs**’ of the model.

**Question 3:** Are there other ways or **other angles of viewing higher education productivity** that you think the model misses?

*## Script: I want to bring the discussion back to practical terms. Other sectors and industries have less trouble talking about ‘waste’ and inefficiency; that is, inefficiency in terms of resources, time, and effort. It is difficult to have this discussion in higher education because of the values surrounding higher education activities.*

**Question 4:** In your experiences, with respect to specific components and operations of higher education institutions, would you be willing to identify key areas **where waste or inefficiency occurs**?

What about the highest performing areas?

**Question 5:** With respect to either reducing inefficiency or adding value/improving quality, if I wanted to optimise institutional performance, what are the key **constraints to optimisation**? I.e., what forces most heavily **shape the status quo**?

## Appendix F

### *Separate Education and Research Productivity Decomposition*

The decomposition requires only one additional non-empirical assumption. The assumption requires a judgement about average time allocations toward research and teaching by staff that do not have sole teaching or research responsibility. The baseline assumption in the current analysis is that staff with both teaching and research responsibility, as well as non-academic staff, provide a 50-50 split of time and effort to be realised as value-add to education and research functions and outputs.

Recall the SFP input variables in Table 31 from Section 7.4.1.

|                         |       |                                |
|-------------------------|-------|--------------------------------|
| Teaching Labour         | $L_3$ | FTEs Teaching Only Staff       |
| Research Labour         | $L_4$ | FTEs Research Only Staff       |
| Teach & Research Labour | $L_5$ | FTEs Teaching & Research Staff |
| Non-Academic Labour     | $L_6$ | FTEs Other Function Staff      |

Consider a the SFP TI calculation using these variables:

$$P_{\Delta t} = \frac{E_{\Delta t}^{vE} \cdot R_{\Delta t}^{vR}}{L_{3\Delta t}^{u_1} \cdot L_{4\Delta t}^{u_2} \cdot L_{5\Delta t}^{u_3} \cdot L_{6\Delta t}^{u_4}} = \frac{E_{\Delta t}^{vE}}{L_{3\Delta t}^{u_1} \cdot \sqrt{L_{5\Delta t}^{u_3} \cdot L_{6\Delta t}^{u_4}}} \cdot \frac{R_{\Delta t}^{vR}}{L_{4\Delta t}^{u_2} \cdot \sqrt{L_{5\Delta t}^{u_3} \cdot L_{6\Delta t}^{u_4}}} \quad [20]$$

Where,

$$u_i = \frac{S_{L_{i,t}} + S_{L_{i,t-1}}}{2}, \quad [21]$$

where  $S_{L_{i,t}} = L_{i,t} / \text{total labour FTEs } (t)$ ,  
and  $S_{L_{i,t-1}} = L_{i,t-1} / \text{total labour FTEs } (t-1)$ .

Thus, separate education productivity change is given by:

$$P_{E\Delta t} = \frac{E_{\Delta t}^{vE}}{L_{3\Delta t}^{u_1} \cdot \sqrt{L_{5\Delta t}^{u_3} \cdot L_{6\Delta t}^{u_4}}} \quad [22]$$

And, separate research productivity change is given by:

$$P_{R\Delta t} = \frac{R_{\Delta t}^{vR}}{L_{4\Delta t}^{u_2} \cdot \sqrt{L_{5\Delta t}^{u_3} \cdot L_{6\Delta t}^{u_4}}} \quad [23]$$

## Appendix G

### *Absolute Productivity Approximation*

First, define a function  $f$  that rescales all data elements in the model to the same units of measure and preserves all variation among observations per data element. Let  $U$  be a data element in the model input-output framework and  $\mathbf{U}$  be the vector of all institutional values for  $U$  in the dataset. And let  $U_i$  be the value of  $U$  for institution  $i$ . Then,

$$U'_i = f(U_i) = \frac{U_i}{\max(\mathbf{U})} \quad [24]$$

Accordingly,  $0 \leq U'_i \leq 1$ . If  $U'_i = 1$ , then institution  $i$  exhibits the largest value for  $U$  across all institutions in the dataset.

So, let  $P_i$  be the absolute productivity of institution  $i$  defined by,

$$P_i = \frac{Y'_i}{X'_i} = \frac{v_E E'_i + v_R R'_i}{X'_i} \quad [25]$$

Where,

$$R'_i = \frac{R'_{l,i} + R'_{p,i} + R'_{r,i}}{3} \quad [26]$$

And where,  $E'_i$  is the scaled value for adjusted load, and  $R'_{l,i}$ ,  $R'_{p,i}$ , and  $R'_{r,i}$  are the scaled values for adjusted load research, adjusted publications, and research income, respectively.

Since productivity is now measured on a new scale, it is helpful to provide a reference point for interpreting results. The feature scaling function implies that a value of  $P = 1$  is equivalent to the maximum observed output in the dataset divided by the maximum observed input in the dataset, even if the input and output values are observed values from different institutions. The explanation is given by the identity shown below, where  $v_E$  and  $v_R$  are the same corresponding education and research value weights from the model's TI productivity change calculation.

$$\begin{aligned} P = 1 &= \frac{\max(Y')}{\max(X')} = \frac{v_E \max(E') + v_R \max(R')}{\max(X')} = \\ &= \frac{v_E f(\max(\mathbf{E})) + v_R f(\max(\mathbf{R}))}{f(\max(\mathbf{X}))} = \\ &= \frac{v_E \frac{\max(\mathbf{E})}{\max(\mathbf{E})} + v_R \frac{\max(\mathbf{R})}{\max(\mathbf{R})}}{\frac{\max(\mathbf{X})}{\max(\mathbf{X})}} = \\ &= \frac{v_E + v_R}{1} = \frac{1}{1} = 1 = P \end{aligned} \quad [27]$$

Recall now from Section 6.4.6 and Appendix B that  $\Pi P_{\Delta t}$  represents the calculation for cumulative productivity change, and that the vector of  $\Pi P_{\Delta t}$  values for institution  $i$  over the period  $t_0$  to  $t_n$  is given by:

$$\mathbf{PP}_{t_0,n}^i = [1, P_{t_{0,1}}^i, P_{t_{0,2}}^i, \dots, P_{t_{0,n}}^i] \quad [28]$$

This vector may also be referred to ‘normative productivity change’ because, by the definition of the TI method, all productivity change indexes are normalised to a value of ‘1’ at the beginning of the period.

Hence, if we wish to find the absolute productivity change over time of institution  $i$  over the period  $t_0$  to  $t_n$  relative to an initial absolute productivity value, we can take our definition of  $P_i$  above and find the value at time  $t_0$ , given by  $P_{i,0}$ . We then multiply by the normative change vector above.

$$P_{i,0} \cdot \mathbf{PP}_{t_0,n}^i = P_{i,0} \cdot [1, P_{t_{0,1}}^i, P_{t_{0,2}}^i, \dots, P_{t_{0,n}}^i] \quad [29]$$

This now gives us ‘relative productivity change’ values over the period with respect to an initial absolute productivity value.

## Appendix H

**Table 46: Institutional absolute productivity indexes 2007**

| Institution                            | edu =<br>0.7 res<br>= 0.3 | edu =<br>0.5 res<br>= 0.5 | edu =<br>0.3 res<br>= 0.7 | edu =<br>0.7 res<br>= 0.3<br>rank | edu =<br>0.5 res<br>= 0.5<br>rank | edu =<br>0.3 res<br>= 0.7<br>rank | Potential<br>Climb in<br>Rank |
|----------------------------------------|---------------------------|---------------------------|---------------------------|-----------------------------------|-----------------------------------|-----------------------------------|-------------------------------|
| Southern Cross University              | 1.47                      | 1.31                      | 1.16                      | 3                                 | 1                                 | 2                                 | 2                             |
| Charles Sturt University               | 1.47                      | 1.28                      | 1.09                      | 2                                 | 2                                 | 3                                 | 1                             |
| University of Southern Queensland      | 1.47                      | 1.27                      | 1.07                      | 1                                 | 3                                 | 5                                 | 4                             |
| University of Wollongong               | 1.32                      | 1.24                      | 1.16                      | 5                                 | 4                                 | 1                                 | 4                             |
| Australian Catholic University         | 1.44                      | 1.24                      | 1.04                      | 4                                 | 5                                 | 10                                | 6                             |
| University of South Australia          | 1.32                      | 1.19                      | 1.07                      | 6                                 | 6                                 | 6                                 | 0                             |
| Macquarie University                   | 1.26                      | 1.17                      | 1.07                      | 9                                 | 7                                 | 4                                 | 5                             |
| Queensland University of Technology    | 1.26                      | 1.16                      | 1.05                      | 10                                | 8                                 | 7                                 | 3                             |
| University of Technology, Sydney       | 1.27                      | 1.15                      | 1.04                      | 8                                 | 9                                 | 11                                | 3                             |
| Curtin University of Technology        | 1.26                      | 1.15                      | 1.04                      | 11                                | 10                                | 9                                 | 2                             |
| University of Western Sydney           | 1.28                      | 1.12                      | 0.96                      | 7                                 | 11                                | 17                                | 10                            |
| Flinders University                    | 1.16                      | 1.09                      | 1.03                      | 16                                | 12                                | 12                                | 4                             |
| Griffith University                    | 1.19                      | 1.08                      | 0.98                      | 13                                | 13                                | 16                                | 3                             |
| University of Tasmania                 | 1.11                      | 1.08                      | 1.04                      | 22                                | 14                                | 8                                 | 14                            |
| Edith Cowan University                 | 1.21                      | 1.06                      | 0.92                      | 12                                | 15                                | 24                                | 12                            |
| The University of Newcastle            | 1.11                      | 1.05                      | 0.98                      | 20                                | 16                                | 15                                | 5                             |
| Deakin University                      | 1.16                      | 1.04                      | 0.92                      | 15                                | 17                                | 23                                | 8                             |
| The University of New England          | 1.13                      | 1.04                      | 0.95                      | 19                                | 18                                | 21                                | 3                             |
| University of Canberra                 | 1.14                      | 1.02                      | 0.90                      | 18                                | 19                                | 27                                | 9                             |
| James Cook University                  | 1.08                      | 1.02                      | 0.95                      | 24                                | 20                                | 20                                | 4                             |
| Central Queensland University          | 1.18                      | 1.01                      | 0.85                      | 14                                | 21                                | 31                                | 17                            |
| Murdoch University                     | 1.06                      | 1.01                      | 0.95                      | 25                                | 22                                | 19                                | 6                             |
| La Trobe University                    | 1.11                      | 1.01                      | 0.91                      | 23                                | 23                                | 26                                | 3                             |
| The University of Sydney               | 0.99                      | 1.01                      | 1.03                      | 29                                | 24                                | 13                                | 16                            |
| University of Adelaide                 | 0.99                      | 1.00                      | 1.01                      | 28                                | 25                                | 14                                | 14                            |
| RMIT University                        | 1.11                      | 1.00                      | 0.89                      | 21                                | 26                                | 29                                | 8                             |
| The University of Notre Dame Australia | 1.14                      | 0.97                      | 0.80                      | 17                                | 27                                | 32                                | 15                            |
| The University of Queensland           | 0.91                      | 0.93                      | 0.96                      | 32                                | 28                                | 18                                | 14                            |

|                                     |      |      |      |    |    |    |   |
|-------------------------------------|------|------|------|----|----|----|---|
| The University of New South Wales   | 0.92 | 0.93 | 0.94 | 31 | 29 | 22 | 9 |
| Monash University                   | 0.96 | 0.93 | 0.90 | 30 | 30 | 28 | 2 |
| University of the Sunshine Coast    | 1.03 | 0.91 | 0.78 | 26 | 31 | 33 | 7 |
| The University of Melbourne         | 0.88 | 0.90 | 0.92 | 34 | 32 | 25 | 9 |
| Federation University               | 1.00 | 0.87 | 0.74 | 27 | 33 | 34 | 7 |
| The University of Western Australia | 0.81 | 0.83 | 0.86 | 36 | 34 | 30 | 6 |
| Victoria University                 | 0.88 | 0.78 | 0.68 | 33 | 35 | 37 | 4 |
| Swinburne University of Technology  | 0.85 | 0.77 | 0.70 | 35 | 36 | 36 | 1 |
| Australian National University      | 0.64 | 0.69 | 0.74 | 37 | 37 | 35 | 2 |
| Charles Darwin University           | 0.56 | 0.55 | 0.54 | 38 | 38 | 38 | 0 |

## Appendix I

### Model Versions Comparison

**Table 47: Model version comparison, system TFP change results for each year 2007-13**

|      | Model V1 |       | Model V2 |       | Model V3 |       |
|------|----------|-------|----------|-------|----------|-------|
| Year | PΔt      | ΠPΔt  | PΔt      | ΠPΔt  | PΔt      | ΠPΔt  |
| 2007 | 1        | 1     | 1        | 1     | 1        | 1     |
| 2008 | 0.962    | 0.962 | 0.960    | 0.960 | 0.969    | 0.969 |
| 2009 | 0.955    | 0.919 | 0.963    | 0.924 | 0.971    | 0.941 |
| 2010 | 0.993    | 0.913 | 0.975    | 0.901 | 0.992    | 0.934 |
| 2011 | 1.011    | 0.923 | 0.972    | 0.876 | 0.986    | 0.920 |
| 2012 | 0.976    | 0.901 | 0.965    | 0.845 | 0.968    | 0.891 |
| 2013 | 0.985    | 0.887 | 0.991    | 0.838 | 0.991    | 0.883 |

**Table 48: Model version comparison, cumulative change results 2007-13**

|                                        | Model V1 |      | Model V2 |      | Model V3 |      |                         |
|----------------------------------------|----------|------|----------|------|----------|------|-------------------------|
| Institution                            | PΔt      | Rank | PΔt      | Rank | PΔt      | Rank | Range of Ranking Change |
| University of Tasmania                 | 0.926    | 10   | 0.831    | 24   | 0.780    | 35   | 25                      |
| The University of Notre Dame Australia | 0.843    | 24   | 0.873    | 14   | 1.155    | 1    | 23                      |
| The University of New England          | 0.788    | 32   | 0.839    | 18   | 0.976    | 11   | 21                      |
| The University of Queensland           | 0.890    | 14   | 0.828    | 25   | 0.839    | 32   | 18                      |
| The University of Sydney               | 0.862    | 18   | 0.802    | 30   | 0.813    | 34   | 16                      |
| Australian Catholic University         | 0.859    | 19   | 0.889    | 10   | 1.050    | 4    | 15                      |
| Deakin University                      | 0.840    | 25   | 0.875    | 13   | 0.991    | 10   | 15                      |
| University of Western Sydney           | 0.761    | 33   | 0.765    | 33   | 0.908    | 18   | 15                      |
| Charles Darwin University              | 0.749    | 34   | 0.755    | 34   | 0.896    | 21   | 13                      |
| Murdoch University                     | 0.961    | 7    | 0.929    | 5    | 0.920    | 16   | 11                      |
| Macquarie University                   | 0.812    | 29   | 0.838    | 19   | 0.866    | 27   | 10                      |
| The University of Melbourne            | 0.844    | 23   | 0.799    | 31   | 0.838    | 33   | 10                      |
| The University of New South Wales      | 0.880    | 16   | 0.833    | 21   | 0.870    | 26   | 10                      |
| University of Southern Queensland      | 0.863    | 17   | 0.833    | 23   | 0.947    | 13   | 10                      |
| Flinders University                    | 0.795    | 31   | 0.813    | 28   | 0.896    | 22   | 9                       |
| Griffith University                    | 0.824    | 28   | 0.833    | 22   | 0.898    | 20   | 8                       |
| James Cook University                  | 0.856    | 20   | 0.850    | 17   | 0.877    | 24   | 7                       |
| The University of Newcastle            | 0.847    | 22   | 0.814    | 27   | 0.855    | 29   | 7                       |
| The University of Western Australia    | 0.948    | 8    | 0.888    | 11   | 0.933    | 15   | 7                       |
| Monash University                      | 0.904    | 11   | 0.857    | 15   | 0.917    | 17   | 6                       |
| University of Technology, Sydney       | 0.834    | 26   | 0.809    | 29   | 0.878    | 23   | 6                       |
| Central Queensland University          | 1.040    | 2    | 1.005    | 3    | 1.023    | 7    | 5                       |

|                                     |       |    |       |    |       |    |   |
|-------------------------------------|-------|----|-------|----|-------|----|---|
| Edith Cowan University              | 0.899 | 12 | 0.892 | 9  | 0.934 | 14 | 5 |
| University of Wollongong            | 0.850 | 21 | 0.837 | 20 | 0.877 | 25 | 5 |
| Australian National University      | 1.010 | 3  | 1.007 | 2  | 1.039 | 6  | 4 |
| Curtin University of Technology     | 0.739 | 35 | 0.726 | 35 | 0.846 | 31 | 4 |
| La Trobe University                 | 0.944 | 9  | 0.896 | 8  | 0.952 | 12 | 4 |
| Queensland University of Technology | 0.833 | 27 | 0.821 | 26 | 0.847 | 30 | 4 |
| University of Adelaide              | 0.804 | 30 | 0.785 | 32 | 0.855 | 28 | 4 |
| University of Canberra              | 0.898 | 13 | 0.876 | 12 | 1.004 | 9  | 4 |
| University of South Australia       | 0.885 | 15 | 0.857 | 16 | 0.907 | 19 | 4 |
| RMIT University                     | 0.988 | 5  | 0.999 | 4  | 1.093 | 2  | 3 |
| Victoria University                 | 1.006 | 4  | 0.910 | 7  | 1.046 | 5  | 3 |
| Swinburne University of Technology  | 1.156 | 1  | 1.080 | 1  | 1.070 | 3  | 2 |
| University of the Sunshine Coast    | 0.974 | 6  | 0.919 | 6  | 1.010 | 8  | 2 |
| Charles Sturt University            | 0.726 | 37 | 0.695 | 37 | 0.753 | 36 | 1 |
| Southern Cross University           | 0.738 | 36 | 0.705 | 36 | 0.753 | 37 | 1 |
| Federation University Ballarat      | 0.722 | 38 | 0.627 | 38 | 0.751 | 38 | 0 |

**Table 49: Real value for money adjustment compared to Moradi-Motlagh et al. (2016)**

| University                             | Model V3.0 FVM |       | DEA Model |       |
|----------------------------------------|----------------|-------|-----------|-------|
|                                        | Rank           | Index | Rank      | Index |
| The University of Notre Dame Australia | 1              | 1.326 | NA        | NA    |
| RMIT University                        | 2              | 1.258 | 9         | 1.252 |
| Swinburne University of Technology     | 3              | 1.231 | 17        | 1.172 |
| Australian Catholic University         | 4              | 1.209 | 24        | 1.116 |
| Victoria University                    | 5              | 1.203 | 11        | 1.224 |
| Australian National University         | 6              | 1.196 | 10        | 1.228 |
| Central Queensland University          | 7              | 1.178 | 25        | 1.116 |
| University of the Sunshine Coast       | 8              | 1.161 | 19        | 1.147 |
| University of Canberra                 | 9              | 1.156 | 4         | 1.325 |
| Deakin University                      | 10             | 1.141 | 16        | 1.181 |
| The University of New England          | 11             | 1.126 | 14        | 1.192 |
| La Trobe University                    | 12             | 1.094 | 13        | 1.219 |
| University of Southern Queensland      | 13             | 1.089 | 27        | 1.094 |
| Edith Cowan University                 | 14             | 1.075 | 21        | 1.127 |
| The University of Western Australia    | 15             | 1.073 | 3         | 1.396 |
| Murdoch University                     | 16             | 1.058 | 1         | 1.513 |
| Monash University                      | 17             | 1.054 | 7         | 1.254 |
| University of Western Sydney           | 18             | 1.046 | 23        | 1.119 |

|                                   |    |       |    |       |
|-----------------------------------|----|-------|----|-------|
| University of South Australia     | 19 | 1.044 | 15 | 1.184 |
| Griffith University               | 20 | 1.033 | 8  | 1.252 |
| Charles Darwin University         | 21 | 1.032 | 35 | 0.988 |
| Flinders University               | 22 | 1.031 | 22 | 1.123 |
| University of Technology, Sydney  | 23 | 1.011 | 26 | 1.1   |
| James Cook University             | 24 | 1.010 | 29 | 1.068 |
| University of Wollongong          | 25 | 1.009 | 32 | 1.041 |
| The University of New South Wales | 26 | 1.000 | 2  | 1.43  |
| Macquarie University              | 27 | 0.997 | 30 | 1.066 |
| The University of Newcastle       | 28 | 0.984 | 20 | 1.144 |
| University of Adelaide            | 29 | 0.984 | 34 | 0.989 |
| Queensland Uni. of Technology     | 30 | 0.974 | 28 | 1.076 |
| Curtin University of Technology   | 31 | 0.973 | 33 | 0.997 |
| The University of Queensland      | 32 | 0.965 | 18 | 1.17  |
| The University of Melbourne       | 33 | 0.964 | 5  | 1.28  |
| The University of Sydney          | 34 | 0.935 | 6  | 1.258 |
| University of Tasmania            | 35 | 0.898 | 12 | 1.22  |
| Charles Sturt University          | 36 | 0.867 | 36 | 0.957 |
| Southern Cross University         | 37 | 0.866 | 37 | 0.957 |
| Federation University             | 38 | 0.861 | 31 | 1.062 |

**Table 50: Institutional rankings' sensitivity to different value weight schemes**

| Institution                            | edu / res<br>0.9 / 0.1<br>Rank | edu / res<br>0.7 / 0.3<br>Rank | edu / res<br>0.5 / 0.5<br>Rank | edu / res<br>0.3 / 0.7<br>Rank | edu / res<br>0.1 / 0.9<br>Rank | Range of<br>Ranking<br>Change |
|----------------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|-------------------------------|
| Murdoch University                     | 4                              | 6                              | 22                             | 34                             | 38                             | 34                            |
| University of Southern Queensland      | 35                             | 28                             | 9                              | 5                              | 3                              | 32                            |
| The University of Melbourne            | 8                              | 15                             | 24                             | 28                             | 31                             | 23                            |
| Central Queensland University          | 26                             | 13                             | 7                              | 4                              | 5                              | 22                            |
| Charles Darwin University              | 6                              | 8                              | 13                             | 24                             | 28                             | 22                            |
| University of Western Sydney           | 29                             | 27                             | 18                             | 11                             | 7                              | 22                            |
| Australian Catholic University         | 27                             | 19                             | 8                              | 6                              | 6                              | 21                            |
| The University of Notre Dame Australia | 20                             | 5                              | 2                              | 1                              | 1                              | 19                            |
| Flinders University                    | 9                              | 18                             | 23                             | 27                             | 26                             | 18                            |

|                                     |    |    |    |    |    |    |
|-------------------------------------|----|----|----|----|----|----|
| The University of Western Australia | 7  | 9  | 17 | 23 | 25 | 18 |
| University of Tasmania              | 15 | 24 | 33 | 32 | 33 | 18 |
| Curtin University of Technology     | 33 | 31 | 25 | 19 | 17 | 16 |
| University of South Australia       | 34 | 34 | 32 | 21 | 18 | 16 |
| The University of Queensland        | 17 | 25 | 31 | 31 | 32 | 15 |
| Griffith University                 | 24 | 20 | 14 | 13 | 10 | 14 |
| University of the Sunshine Coast    | 16 | 7  | 3  | 2  | 2  | 14 |
| Australian National University      | 3  | 3  | 5  | 9  | 14 | 11 |
| The University of New South Wales   | 19 | 23 | 27 | 30 | 29 | 11 |
| The University of Sydney            | 25 | 33 | 35 | 36 | 36 | 11 |
| Monash University                   | 10 | 14 | 15 | 17 | 20 | 10 |
| Queensland University of Technology | 31 | 29 | 26 | 22 | 21 | 10 |
| University of Technology, Sydney    | 23 | 21 | 19 | 15 | 13 | 10 |
| Swinburne University of Technology  | 2  | 2  | 4  | 7  | 11 | 9  |
| The University of Newcastle         | 32 | 32 | 30 | 25 | 23 | 9  |
| University of Wollongong            | 13 | 16 | 20 | 20 | 22 | 9  |
| University of Adelaide              | 21 | 26 | 28 | 29 | 27 | 8  |
| Federation University               | 36 | 36 | 36 | 33 | 30 | 6  |
| Macquarie University                | 28 | 30 | 29 | 26 | 24 | 6  |
| Deakin University                   | 14 | 12 | 10 | 10 | 9  | 5  |
| James Cook University               | 30 | 35 | 34 | 35 | 34 | 5  |
| La Trobe University                 | 11 | 10 | 12 | 14 | 15 | 5  |
| Edith Cowan University              | 22 | 22 | 21 | 18 | 19 | 4  |
| RMIT University                     | 5  | 4  | 6  | 8  | 8  | 4  |
| Southern Cross University           | 38 | 38 | 37 | 37 | 35 | 3  |
| Victoria University                 | 1  | 1  | 1  | 3  | 4  | 3  |
| The University of New England       | 18 | 17 | 16 | 16 | 16 | 2  |
| Charles Sturt University            | 37 | 37 | 38 | 38 | 37 | 1  |
| University of Canberra              | 12 | 11 | 11 | 12 | 12 | 1  |