Measuring International Higher Education Productivity: Lessons from nine countries in Asia

Kenneth Moore*, Hamish Coates** and Gwilym Croucher***

Abstract. The paper offers a synthesis of findings from an international initiative coordinated by the Asia Productivity Organisation (APO) to measure higher education productivity across nine Asian countries. The paper discusses benefits, barriers, and potential for estimating university productivity across international contexts. Stakeholders from nine participating countries collected and analysed institutional data and reported measurement results. The APO initiative represents the first multi-country test of an adapted productivity measurement model first advanced by the United States National Research Council (NRC). The research provides evidence for proof of concept of the adapted NRC model for use across international contexts. Additional findings demonstrate the range of productivity definitions and interpretations for higher education. The paper concludes by showing priority areas for both targeted and broad developments in research and practice of measuring productivity in higher education.

Keywords: higher education, productivity, measurement, Asia

Introduction

The purpose of the paper is to share findings that can help higher education stakeholders in Asia extract more value from institutional productivity estimation. Findings are intended to guide the practice of higher education productivity measurement in Asia by illustrating how to develop more robust measurement tools and how the tools can be better applied in context. The study focuses on the experiences of higher education stakeholders who measured the productivity of hundreds of higher education institutions across nine Asian countries. The stakeholders participated in an international initiative coordinated by the Asia Productivity Organisation (APO). The APO initiative was designed as a scoping study to identify the potential and the limitations of measuring higher education productivity in Asia. Full details are available in Coates (2017a). The current paper synthesises

*Researcher, University of Melbourne, Australia, E-mail: kennethm1@student.unimelb.edu.au
** Professor, Tsinghua University, China, E-mail: hamishcoates@tsinghua.edu.cn
*** Senior Lecturer, University of Melbourne, Australia, E-mail: gscrou@unimelb.edu.au

findings from this initiative and highlights priority areas for improving productivity measurement in Asia.

The purpose of the current study is not to provide a comparison of empirical trends across countries or to make authoritative statements about institutional productivity in Asia. The study of higher education productivity is not yet well-established. Definitions and understandings of university productivity and performance are neither precise nor universal across the higher education landscape (Massy, 2016; Coates, 2017b). Different countries maintain different priorities for higher education reform and development, and they face different challenges to their systems. A commonality between countries, however, is public concern about higher education performance. Performance issues relate to higher education funding, social and cultural impact, contribution to innovation, and human capital formation (OECD, 2017a). The ubiquity of scholarly discourse regarding such concerns may date back to the transition from elite to mass higher education, and this concern has continued into the 21st century (Trow, 1973; Alexander, 2000). The demand for improving performance in higher education has increased the demand for better information on higher education systems (OECD, 2017b). The current study provides evidence for how higher education stakeholders interested in measuring institutional productivity can reform measurement efforts in context to achieve results with increased value and utility for decision-making.

Context

Recent years have seen increased interest in both the measurement and the improvement of measurement of higher education productivity (Sullivan, Massy, Mackie & Sinha, 2012; Lee & Worthington, 2016; Massy & Archer, 2018; Moore, Coates & Croucher, 2018a). Institutional productivity encompasses the efficiency and effectiveness of the key work processes of institutions (Miller, 2007). Productivity measurements are taken using performance indicators that serve as proxies for institutional inputs and outputs. Many quantitative techniques exist for calculating productivity estimates, but all techniques produce output to input ratios (Bairam, 1994).

The National Research Council (NRC) of the United States National Academy of Sciences developed a productivity measurement technique specific for higher education (Sullivan et al., 2012). The technique is henceforth referred to as the NRC model, and it uses Törnqvist indexing as the methodological approach for aggregating data and estimating productivity. The Törnqvist index measures total factor productivity and productivity change. Respectively, the index represents productivity that accounts for all factors of production and for rates of growth or decline in productivity (Sullivan et al., 2012). This type of productivity measurement has proven useful and reliable for numerous industries and is espoused by the OECD (2001) and the United States Bureau of Labor Statistics (BLS, 2007). The index's widespread use in productivity measurement can be attributed to its accessibility and to its designation as a "superlative index", as shown by Caves et al.

(1982). The technique differs from alternative methods for calculating university productivity, such as data envelopment analysis or stochastic frontier analysis, as no advanced mathematical programming or econometric techniques are required for the Törnqvist index. The index calculates percentage change ratios from data elements present in the model from one time point to the next.

Moore, Coates and Croucher (2018b) provide mathematical specifications for how the NRC model can be adapted and generalised for use in international contexts. A prime contribution of the original NRC model is its specification of an education output indicator termed "adjusted credit hours" that accounts for both the scale of educational services that an institution provides and the additional value that credit hours have when they are accumulated and organized into a completed degree (Sullivan et al., 2012). "Credit hours" is a measure of course load specific to the United States higher education system, but Moore et al. (2018b) show how the calculation of "adjusted credit hours" in the adapted NRC model can be altered using a generic indicator of equivalent full-time student load to better suit international contexts and other countries' systems that have different methods for determining student load within institutions. The adapted calculation is referred to as "adjusted load". The adapted NRC model and its adjusted calculation techniques were advanced by the APO for use in their nine-country study of higher education productivity in Asia (Coates, 2017a).

Understanding and improving the productivity of higher education is of growing importance in Asia. University education and research is playing an increasingly significant role in Asia (UNESCO, 2014), as is Asia to higher education globally. Between 1980 and 2050 the "center of gravity" of higher education—as with the global economy—is shifting east. Asia now has more higher education students than any other continent and growing research impact (Quah, 2011). Scaling higher education while improving quality remains a top priority.

Successful higher education system scaling and improvement depends upon the availability of accurate and reliable information on areas of interest (Sullivan et al., 2012). From a global perspective, higher education systems may appear to evolve organically alongside tectonic shifts in the landscape. Changes to higher education structures, operations, funding, and even academic work have often been viewed as consequences of higher education expansion (Trow, 1973; Clark 1983; Schofer & Meyer, 2005). Higher education expansion, for instance, has often been associated with broad changes to resource management, academic work, and even the purpose of higher education. Many other dynamics and forces are also at work, however, including new media and technologies, new management systems, and the growth of knowledge societies (Teichler, 2006). Enacting positive change amidst such complexity requires nuanced understandings and definitions of problems. Reforming and changing systems in specific settings, then, depends upon the availability of clear and relevant information about higher education problems in context.

Generating accurate and reliable information on higher education institutions, however, is difficult. Perceptions of higher education quality, for instance, may vary from individual to individual. Coates (2017b) notes that higher quality in higher education is difficult to assess because

of its nature as a "credence good". Credence goods are defined as goods or services where the seller or provider knows more about their quality than the consumer, and thus, quality cannot be judged until after consumption (Dulleck & Kerschbamer, 2006). The elusive nature of higher education quality drives some of the most important considerations about the accuracy of performance and productivity metrics in higher education. Measurement limitations associated with less tangible higher education variables, however, do not preclude the ability to extract useful information from other types of institutional variables.

Improving information on the operational dimension of higher education performance is critical where governments and societies are taking active steps to reform higher education systems. The study of higher education productivity examines tangible inputs and outputs of institutions to represent the functions, operations, and key work processes that institutions perform (Miller, 2007). Productivity estimates speak to both institutional efficiency and effectiveness. Productivity estimates may serve as lead indicators for longer term success. They can help stakeholders understand whether institutions are operating in accordance with stated mission objectives and whether they are poised to deliver on their promised outcomes. Productivity does not encompass the full range of higher education performance. Direct representations of quality, outcomes, and impact still fall outside the scope of institutional productivity estimates (Massy & Archer, 2018). As lead indicators of performance and of performance trajectories, however, productivity estimates hold unique potential for generating information to enhance decision-making in emerging Asian higher education systems, where tests and implementations of new and hybrid models for policy, operations, and governance are common (Huang, 2007; Marginson, 2011).

Conceptual framework

The current research uses design thinking to synthesise findings from the APO initiative. Participants of the APO initiative tested the adapted NRC measurement model from Moore et al. (2018b) for estimating higher education institutional productivity. The current research generates findings that can advance the study higher education productivity in Asian systems by signalling what can be further improved about the measurement tool and how it can be better applied in context. Design thinking operates under the assumption that humans are rarely able to identify all possible solutions to problems and therefore settle for choices that satisfy desired solution properties (Rowe, 1991). Design thinking facilitates value creation under circumstances where both specific solutions and optimal key work processes are unknown (Dorst, 2011). Extracting more value from productivity estimates under variable and complex institutional conditions in Asia is the objective of this research.

A design thinking frame was also chosen because it holds the potential to unite the technical and contextual aspects of higher education productivity research. Productivity is a mathematical concept, but institutional data is not context-free. Design thinking creates a bridge between the scientific,

context-independent aspects of a problem and the qualitative, context-dependent aspects of a problem (Plattner et al., 2016). Kelley (2013) explains that a core element of design thinking is the systematic integration of stakeholder experiences and views on problems of interest, so that relevant human contexts may be addressed. Applying design thinking to productivity measurement carried out in the APO initiative can inform (a) the technical aspects of developing a robust university productivity measurement model, and (b) the contextual factors and socio-political dimensions of using and interpreting higher education data.

Method

The objects of analysis of the current research are the results and the implementation reports generated from the APO initiative in Coates (2017a). 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. The APO engaged member countries and selected national experts to participate in the research (Coates, 2017a). Key criteria for participant selection included experience with quantitative research and individuals' access and permission to use institutional data from their respective countries. Not all participants were econometricians or modelling experts, however, so each was provided guidelines for using the measurement tool and for analysing and reporting results.

The APO instructed participants to report on: (a) national interests and intentions for measuring higher education productivity; (b) the datasets they used; (c) implementation of the measurement tool; (d) empirical measurement results; and (e) intentions for advancing higher education productivity measurement in their own countries. Each section of all reports is assessed in the research to generate information for how the practice of productivity measurement can be improved. Operationally, this means producing evidence for how the adapted NRC model can be improved and for how it can be better applied in context. Both technical and contextual information is thus needed for an effective model assessment. Model evaluation under design science research operates under the assessment of four key criteria that target both technical and contextual aspects of the model. The criteria include: (1) model constructs; (2) construct associations; (3) model boundaries; and (4) novelty, revelation, and importance (Williamson & Johanson, 2017).

Assessing model constructs and construct associations involves checking for rigorous definitions and identifying relationships between the constructs and their real-world instances (Williamson & Johanson, 2017). It involves investigating the nature and appropriateness of the associations between constructs. The assessment is performed by examining all measurement model specifications given in the country reports. Assessing model boundaries involves determining whether their scope and limitations are clear. It involves determining what the model represents and does not represent. The assessment is carried out by checking the nature of all countries' measurement results and how results

relate to the countries' measurement contexts and intentions. Finally, assessing novelty, revelation, and importance involves determining what the model illuminates that was not previously known or well-understood. The assessment is performed through synthesising the conclusions that participants made from their measurement results and how the conclusions relate to the counties' intentions to advance higher education productivity measurement in their local contexts.

Table 1 describes input and output indicators and potential data elements for the adapted NRC model. Input indicators include monetary values for labor, capital and intermediaries to account for total factor productivity. Labor is defined as all direct employee expenses. Capital includes an institution's non-current assets, that is, assets from which value is extracted for longer than a single fiscal year. Intermediaries include operational, administrative and non-labor expenditures within a single fiscal year. The input data elements represent common budget categories but are not intended to be prescriptive. They are aggregated to account for the total sum of yearly operational expenses, as well as all non-current assets used during operations. However, actual data elements will vary across countries and institutions depending on measurement and accounting practices, as well as interests in higher education performance.

Table 1. Indicators and data elements[1]

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)
		Administration and other expenses (I_2)
Education outputs	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)
	Learning Outcomes	Proportion of learning outcomes achieved (E_4)
Research outputs	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)

[1] Full mathematical specifications of the adapted NRC model are available in Moore et al. (2018b).

Four potential education outputs are included. Student load and coursework completions allow for calculation of "adjusted load", and graduate employment and learning outcomes are provided as options for systems that prioritise educational outcomes. Based on prior work (Coates, 2016) five potential research outputs are included as measurement options. They are publications, citations, patents, research completions, and research funds. Research funds is intended as an output indicator for systems that award research funding based upon competitive grant applications. Research grants may be viewed as outcomes of a successful grant application.

Results

Country highlights

Each country collected data under different conditions. Table 2 summarises each country's institutional samples from which productivity data was gathered and the methods used to collect the data. The sample sizes vary from country to country depending on the size of the country, the availability of data, and the capacity of the in-country research team. Fijian stakeholders, for instance, gathered data from two of the three universities in the country. Indonesian stakeholders analysed a larger number of institutions, but their sample includes less than ten per cent of the country's over 2000 higher education institutions. Different types of institutions also comprised the different countries' samples. The four institutions in Cambodia's sample are universities. India's sample includes technical institutions. Finally, study participants used three prime strategies for collecting data. Some participants gathered data via survey instruments that required institutional representatives to provide time-series data on the indicators from Table 1 above. 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.

The different national contexts influenced the type of institutional data available and methods for productivity measurement. Table 3 shows the different combinations of education and research indicators collected and measured by each country. The table also summarises the measurement technique employed by each country. Seven of the nine countries used Törnqvist indexing with the adapted NRC model advanced in the study. Two countries cited reasons for not using the technique. Malaysian stakeholders already had expertise using a different technique, the Laspeyres index, also commonly used for price and cost indexing (Braithwait, 1980). Stakeholders in the Philippines cited a different reason. They 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 productivity change indexes.

Table 2. Country samples and data collection

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

Table 3. Academic output indicators and methods employed by country

		Adapted NRC model productivity calculation							Other technique	
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 countries' data selections have a direct bearing on the range of definitions and interpretations for productivity across the higher education landscape. Surface-level implications of the indicators used in the model may guide interpretation of measurement results. Each indicator's inclusion or exclusion in the model has deeper meaning. The indicators chosen for measurement in each context

speak to the interests and priorities of the individual countries and the aspects of higher education that matter to their systems and societies. Data collection, storage, and reporting are time and energy consuming endeavours, and accessibility speaks to interests. The indicators listed in this paper are not exhaustive, and it is no surprise that different systems and different countries define and collect performance indicators in different ways. The results from Table 3 give a sense of the scope of different combinations of data interpretations that may constitute a portrayal of institutional productivity.

Evaluation of measurement model application

Table 4. Findings from the measurement model evaluation

Evaluation Criteria	Key findings from criterion assessment
Novelty, revelation and importance	• Country reports provide evidence for proof of concept of the adapted NRC model for operating on diverse datasets and in a variety of international contexts. • Results indicate the efficacy of productivity change estimates for generating unique and important information for decision-making within dynamic policy and funding environments and within expanding higher education systems. • Measurement results revealed or intensified a sense of urgency among participants to further develop higher education productivity measures.
Model boundaries	• The model has limited ability to reliably represent fluctuations in year-on-year productivity change. • A second limitation in the scope of the model is its ability to account for quality of inputs and outputs.
Model constructs	• The key constructs of “total factor productivity” and “productivity change” proved useful and appropriate for international higher education contexts. • Use and selection of constructs speaks to the need for added rigor in defining model elements. Added precision for “citations”, “graduate employment”, and “learning outcomes” would increase the reliability of productivity estimation.
Construct associations	• The Törnqvist index proved clear and effective for aggregating multiple data elements and for providing interpretable results. • The ‘adjusted load’ education output indicator from the adapted NRC model needs further refinement and unique specification for individual country contexts.

The evaluation of measurement model application produced nine key findings for future directions in higher education productivity measurement in Asia. Findings range from insights about how to refine productivity measurement tools for various international contexts to insights about what phenomena higher education productivity change models are best suited to inform. Findings speak to

the potential and the limitations of measuring higher education productivity. Evaluation results were generated through assessing the four design science criteria described in the preceding section in relation to the individual sections of each country report from Coates (2017a). Table 4 lists key findings. Implications are discussed in the following section.

Discussion

The following discussion centers on implications of research findings for improving the technique and the practice of higher education productivity measurement in Asia. Variability among participants' institutional sample sizes, data collection methods, and types of data analysed means that empirical results reported by each country cannot be extrapolated or used to make inferences about precise magnitudes of higher education productivity across countries or between countries. Rather, findings from this research confirm that the study of higher education productivity remains in its infancy. Productivity metrics still need work before offering valid and reliable indicators of institutional performance, but findings from this research confirm the potential in reforming measurement practice. Lessons learned from research on the APO exercise can further be applied to advance the study of global higher education productivity more broadly.

The different input and output combinations chosen by each country in the study also have a bearing on productivity studies for single institutions within systems. Individual institutions serving different students and operating under different mission objectives also have better and worse input and output combinations for portraying productivity. The results in Table 3 above illustrate the importance of entertaining multiple different portrayals to tell more complete and comprehensive stories about the functioning of institutions and systems. This does not mean searching for the most complimentary portrayal among different options. It means considering a limited number of viable productivity portrayals within clear definitional confines according to both theory and context.

Findings from the assessment of the measurement model's novelty, revelation and importance help to frame the rest of the findings from the research. The assessment helps to illustrate what contexts and definitions may be appropriate for consideration in defining and interpreting productivity metrics. 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 participation in the study 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 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 (Coates, 2017a). Each of these contextual issues serves as a candidate for how productivity metrics can be both framed prior to

analysis and interpreted after analysis.

The successful application of the adapted NRC model across diverse contexts serves as evidence for proof of concept that the model can provide results with utility and relevance across higher education contexts. Findings illustrate that estimates of "total factor productivity" and "productivity change" with Törnqvist indexes generate valuable information from often disparate data elements. In the initial and exploratory stages of studying higher education productivity, the fact the adapted NRC model also requires no special software and incorporates no regression analysis is a positive. Regression analysis is best suited for understanding the strength of relationships, for prediction, and for drawing insight about cause and effect. Countries with different interests in measurement, unreliable data, and unsettled definitions of the phenomena being measured need not yet be concerned with hypothesis testing, statistical significance, or residuals in linear or non-linear model fitting. The exercise has revealed that Törnqvist estimates of productivity change add sufficient precision to conversations about what aspects of performance are important in context and about how performance measurement can be improved without offering premature judgements about empirical rules for production in higher education. Measuring productivity change with Törnqvist indexes limits the ability to compare productivity results between countries and institutions. Only rates of change can be compared using the method. As discussed above, however, only after considerable groundwork and setting clear boundaries could one institution or one country's productivity be compared to another's. The limitation for Törnqvist indexes is a strength at this stage because it encourages deeper analysis of single countries and institutions before attempts to benchmark between institutions and systems.

Findings from the model boundaries assessment showed where future efforts to improve productivity measurement should take place. The prime intention of each country's report was to portray productivity change over a designated period. Country results from the adapted NRC model frequently showed sharp spikes and rebounds in productivity change (Coates, 2017a). The inconsistency of results calls into question the ability of the model to capture true productivity change trends. Universities are often associated with idiosyncratic approaches and a commitment to tradition that can lead to delays between policy implementation, practice, and results (Teichler, 2006; Kezar, 2011). When measured productivity change shows sharp fluctuations from year to year, questions must be raised about the reliability of the measurement tool.

Source data from the countries explains fluctuations in productivity change trends and offers a solution. Budgets for public institutions in the countries' samples had less to do with internal institutional conditions than with changing public policy and ministerial dynamics (Coates, 2017a). Productivity metrics should be designed to take this into account. The prime outputs of higher education institutions 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 assessment of model boundaries also highlighted the difficulty of the model to capture aspects of quality. This is not a surprise. Productivity is an inherently mathematical and quantitative concept. However, higher education productivity estimates should incorporate the value of outputs (Moore, 2018). The NRC's development of the "adjusted credit hour" metric is a direct attempt to incorporate value within a university education productivity estimate, which goes beyond solely quantitative representations of student throughput. The examination of the model construct associations found, however, that "adjusted credit hour" or "adjusted load" indicators were difficult to calculate on a country-by-country basis because of the different types of data available. "Adjusted load" calculations can be useful when adapted to specific contexts, as in Moore et al. (2018a). When different countries calculate student load differently and when program lengths are variable, however, calculations suffer in their validity and reliability. The extent to which productivity metrics should account for output value—and whether quality analyses are best conducted solely in parallel to productivity analyses—remains an open question for continued research.

The model construct assessment speaks further to the issue of representing value and quality in productivity estimates. The results from the previous section summarise all input and output indicators considered by each country for their productivity estimates. Interpretations of academic productivity were very different from country to country, and each had a different outlook and intention for capturing the value and quality of outputs. Specifically, the APO did not provide precise guidelines for representing the output constructs of "graduate employment", "learning outcomes" and "citations". The lack of clear definitions was intentional to generate insight about the differences in what different countries found important and what was available for assessment. Future international productivity measurement initiatives would need to provide specific definitions of all indicators if empirical results were to be compared or used for decision-making. The flexibility allowed for exploring phenomena important to individual contexts and to lay foundations for meaningful country-specific developments.

Conclusion

Findings from the research illustrate interests, capabilities, and intentions for measuring higher education productivity across systems in Asia. Myriad rationales exist to care about the nature and improvement of higher education productivity and its measurement, including:

- Understanding scale economies in higher education.
- Determining implications of policy and regulations.

- Merging or splitting institutions.
- Testing program pricing scenarios.
- Justifying cross-subsidization of programs.
- Expanding, contracting or eliminating programs.
- Framing results of traditional and novel organizational approaches.
- Framing results of idiosyncratic and cookie-cutter organizational approaches.

The state of international higher education productivity measurement is still far from being able to produce reliable and valid estimates for cross-country analysis. Making calculations on existing data is always possible, but the multi-dimensional nature of higher education means that many different types of academic productivity exist for institutions. A priority for future development of the study of international higher education productivity in Asia is organising and validating the range of data options available for calculating estimates in context. No single definition of higher education productivity will ever suit all analytical purposes. As raised in Moore et al. (2018a), different definitions of productivity and different value judgements on individual data elements in calculations can impact both the magnitude of measured results and the interpretation of results. More definitive estimates of institutional productivity may be possible after further advancement on the contextual side of higher education productivity measurement. Applying sophisticated mathematical and econometric techniques on higher education data may be tempting, but real progress toward accuracy and precision in estimation will come with definitional and contextual work on concepts and data used in calculations.

With attrition of ill-defined and roughly understood higher education concepts and phenomena, countries in Asia may be able to collect and report data with more reliability and gain progress toward meaningful and targeted benchmarking. Participants of this study reported their own measurement results with both excitement and skepticism. Only well-defined and well-conceived metrics can provide baselines for understanding performance and sparking conversation. Higher education institutions will always have both physical and ideological components that need attention. Productivity estimates serve the purpose of helping to inform how resource management and operational practices and constraints either hinder or facilitate the success of higher education's core functions. Indicators for higher education, though, are always tied to interests, intentions, and ideologies.

Acknowledgements

The authors acknowledge the support of the Asia Productivity Organisation (APO) in coordinating and leading the project titled "Research on Measuring Productivity in Higher Education" in which Hamish Coates served as Chief Investigator. We are grateful for the APO's leadership and for the contribution

of national experts and all others who contributed to the study.

References

Alexander, F.K. (2000). The changing face of accountability: Monitoring and assessing institutional performance in higher education. *The Journal of Higher Education*, *71*(4), 411-431.

Bairam, E. (1994). *Homogeneous and Nonhomogeneous Production Functions: Theory and applications.* Brookfield: Ashgate.

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.

Bureau of Labor Statistics (BLS). (2007). *Technical Information about the BLS Multifactor Productivity Measures.* Retrieved August 16, 2016, from www.bls.gov/mfp/mprtech.pdf

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.

Clark, B. R. (1983). *The higher education system : academic organization in cross-national perspective*. Berkeley, Calif. : University of California Press, c1983.

Coates, H. (2016). Performance measurements and alternatives to rankings. 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.* Abingdon: Taylor and Francis.

Coates, H. (Ed.). (2017a). *Productivity in higher education: Research insights for universities and governments in Asia.* Tokyo: Asian Productivity Organisation.

Coates, H. (2017b). *The market for learning: Leading transparent higher education*. Dordrecht: Springer.

Dorst, K. (2011). The core of 'design thinking' and its application. *Design studies*, *32*(6), 521-532.

Dulleck, U., & Kerschbamer, R. (2006). On doctors, mechanics, and computer specialists: The economics of credence goods. *Journal of Economic Literature*, *44*(March), 5–42

Huang, F. (2007). Internationalization of higher education in the developing and emerging countries: A focus on transnational higher education in Asia. *Journal of Studies in International Education*, *11*(3-4), 421-432.

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.

Kelley, T. & Kelley, D. (2013). *Creative confidence: Unleashing the creative potential within us all*. New York: Crown Business.

Kezar, A. (2011). *Understanding and facilitating organizational change in the 21st century: Recent research and conceptualizations: ASHE-ERIC Higher Education Report, Volume 28, Number 4* (Vol. 155). John Wiley & Sons.

Marginson, S. (2011). Higher education in East Asia and Singapore: Rise of the Confucian model. *Higher Education*, *61*(5), 587-611.

Massy, W.F. (2016). *Reengineering the university: How to be mission centered, market smart and margin conscious.* Baltimore: Johns Hopkins University Press.

Massy, W.F. & Archer, S. (2018). Perspectives and advances on productivity measurement in higher education. In Hazelkorn, E., Coates, H. & McCormick, A. (Eds.). *Research handbook on quality, performance and accountability*. Cheltenham: Edward Elgar.

Massy, W.F., Sullivan, T.A., & Mackie, C. (2013). Improving measurement of productivity in higher education. *Change: The magazine of higher learning, 45*(1), 15-23.

Miller, B.A. (2007). *Assessing organizational performance in higher education*. Jossey-Bass.

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. *Research handbook on quality, performance and accountability in higher education*, 161.

Organization for Economic Cooperation and Development (OECD). (2001). *Measuring productivity: Measurement of aggregate and industry-level productivity growth.* Paris: OECD.

Organization for Economic Cooperation and Development (OECD). (2017a). *Education at a glance 2017.* Paris: OECD.

Organization for Economic Cooperation and Development (OECD). (2017b). *Benchmarking higher education system performance: Conceptual framework and data.* Enhancing Higher Education System Performance. Paris: OECD.

Plattner, H., Meinel, C., & Leifer, L. (Eds.). (2016). *Design thinking research: Making design thinking foundational*. Springer, Cham.

Quah, D. (2011). The global economy's shifting centre of gravity. *Global Policy, 2*, 3-9.

Rowe, P.G. (1991). *Design thinking*. MIT press.

Schofer, E., & Meyer, J.W. (2005). The worldwide expansion of higher education in the twentieth century. *American Sociological Review*, *70*(6), 898-920.

Sullivan, T.A., Mackie, C., Massy, W.F., & Sinha, E. (2012). *Improving measurement of productivity in higher education.* Washington DC: National Academies Press.

Teichler, U. (2006). Changing structures of the higher education systems: The increasing complexity of underlying forces. *Higher Education Policy*, *19*(4), 447-461.

Trow, M. (1973). Problems in the Transition from Elite to Mass Higher Education. Berkeley, CA.

United Nations Educational, Scientific and Cultural Organisation (UNESCO). (2014). Higher Education in Asia: Expanding out, Expanding up. UNESCO Institute for Statistics. Montreal, Quebec

Williamson, K., & Johanson, G. (Eds.). (2017). *Research methods: Information, systems, and contexts*. Chandos Publishing.