Value-Added: How do you measure whether universities are delivering for their students? HEPI 2015 Annual Lecture

Andreas Schleicher

Higher Education Policy Institute

HEPI Report 82

About the Author

Andreas Schleicher is Director for Education and Skills, and Special Advisor on Education Policy to the Secretary-General at the Organisation for Economic Co-operation and Development (OECD) in Paris.

As a key member of the OECD Senior Management team, Mr. Schleicher supports the Secretary-General's strategy to produce analysis and policy advice that advances economic growth and social progress. He promotes the work of the Directorate for Education and Skills on a global stage and fosters co-operation both within and outside the OECD. In addition to policy and country reviews, the work of the Directorate includes the Programme for International Student Assessment (PISA), the OECD Survey of Adult Skills (PIAAC), the OECD Teaching and Learning International Survey (TALIS), and the development and analysis of benchmarks on the performance of education systems (INES).

Before joining the OECD, Mr. Schleicher was Director for Analysis at the International Association for Educational Achievement (IEA). He studied Physics in Germany and received a degree in Mathematics and Statistics in Australia. He is the recipient of numerous honours and awards, including the 'Theodor Heuss' prize, awarded in the name of the first president of the Federal Republic of Germany for 'exemplary democratic engagement'. He holds an honorary Professorship at the University of Heidelberg.

Foreword

Professor Sir Ivor Crewe, Chair of HEPI's Trustees

The Annual Lecture is the biggest event in HEPI's calendar and the 2015 event was the 12th in the organisation's history. As in earlier years, it would not have been possible to organise the Annual Lecture without the generous support of two of HEPI's longstanding sponsors, Pearson and Wiley, to whom we are most grateful.

This year's speaker was Andreas Schleicher, the Director of Education and Skills at the Organisation of Economic Co-operation and Development (OECD) based in Paris. His background is in Physics and Mathematics but his career has been devoted to research on educational achievement, most notably perhaps the development of the Programme for International Student Assessment (PISA) study.

Andreas Schleicher has unrivalled knowledge of education systems across the globe, which is displayed in the annual *Education at a Glance* publication, which weighs in at a hefty 568 pages this year. It is an unrivalled source of comparative evidence.

As a result of Andreas' expertise and reputation, UK politicians on all sides of the political spectrum regularly pray him in aid. The press coverage of *Education at a Glance 2015* dwelled on its revelation that England has the highest average tuition fees among publicly-funded universities in the OECD but also on Andreas' comment that the English system is fair, efficient and

likely to be sustainable. No wonder one well-known higher education policy expert, Andy Westwood, recently asked, 'The OECD, whose side are they on?'. The answer is: on the side of robust, comparative and openly-available evidence, carefully and impartially interpreted.

Value-Added: How do you measure whether universities are delivering for their students?

Andreas Schleicher

The 2015 HEPI Annual Lecture was delivered on 1st December 2015 in central London. The slides that accompanied the lecture are freely available via http://www.hepi.ac.uk/2015/12/01/2015-annual-lecture-by-andreas-schleicher-director-of-education-at-the-oecd/.

Introduction

It is a great privilege and pleasure to address you. The theme you have given me – measuring whether universities are delivering for their students – is one I am passionate about. We owe our students an honest answer to this question; we owe it to the parents who pay a lot of money towards higher education; and we owe it to employers who recruit graduates. I am not only talking about the money that goes into higher education, as the most valuable resource is probably the time students spend in universities.

There are three reasons why we should look at the measurement question from an international angle. First, there is internationalisation: higher education has become a global endeavour.

The second is the need to learn from diversity. Some people say, 'If you develop metrics internationally, it is going to lead to standardisation and uniformity.' I believe the opposite. By

looking at ourselves in the mirror of what others are achieving, we learn much about ourselves and how we can respond to common challenges differently. That is the experience we have had with PISA. When we launched *Education at a Glance* in a UK school last year, there were a group of teachers next door from Shanghai, discussing with British teachers how to teach Mathematics. Would that have happened without the comparative framework of PISA that allows us to see what is possible and to explore how people across the world think differently and work differently?

The third aspect is associated with capacity. The recent higher education green paper, *Fulfilling our potential: Teaching excellence, social mobility and student choice*, from the Westminster Government encourages the measurement of higher education learning outcomes, but why has no country made anything like this happen yet? Part of the answer is that it is very difficult and complicated and by pooling expertise and experience internationally we can go much further – and faster.

Internationalisation

Let me start with internationalisation. When you look at student enrolment, you can see that China and India were big players in 2013 already, but you see that by 2030 they will be far more prominent. By 2030, we may be seeing 40% of all STEM graduates coming from China alone.

Do Asian universities have a fair chance to compete on any of the metrics that we are currently using to judge their success?

Certainly not. If we look at past reputation rather than current outcomes as our main determinant for the relative standing of universities, no entrant can compete with the incumbents.

Signalling

Another aspect is important too: degrees and qualifications are signals. But how good are those signals in revealing what people know and what they can do with what they know?

Our survey of Adult Skills revealed that in Italy, those with university qualifications tended to have higher numeracy skills than those with school qualifications. But there is a surprising amount of overlap. Some of the school leavers turn out to be quite highly skilled and some of the university graduates are not so well skilled. There are a lot of things going on in the data: it is a cross-sectional picture, with some people continuing to learn throughout their lives and others losing skills. Yet it shows that degrees and qualifications are not always a good predictor of the skills that people currently have. Those differences become even more pronounced when we look at this across countries: it turns out that Japanese high-school graduates come out better than Italian university graduates on foundation skills like literacy and numeracy.

Had we measured other skills, we might have got a different picture. But it goes to show that there are things to learn from becoming better at measuring skills and knowledge rather than just looking at degrees that may have the same names but not necessarily the same content across countries.

Learning outcomes as the key

All this raises the question of whether we can do better in assessing the value that universities create, by measuring student learning outcomes directly. To address this, let me try to address three questions:

1. Why measure learning gain internationally?
2. How can we measure learning gain?
3. How do we know that our measures actually reflect the quality of higher education learning outcomes?

Why?

There was a time when people looked to universities to judge the quality of education. Today, it is the other way around: the public want better information on the quality of universities.

Over the last thirty years, the focus of higher education has changed significantly, primarily in response to the changing nature of work. A rapid increase in jobs requiring higher-order cognitive skills has created a worldwide need for more graduate employees. As a result, the priority previously given by universities to inducting a small minority into research capabilities has given way in many countries to providing up to half the population with the skills and knowledge relevant to employability. This has been achieved through a rapid expansion of the higher education sector and the establishment of more diverse types of higher education institutions.

This historic shift has been accelerated by changes in funding regimes. The rising costs of higher education are increasingly borne by students themselves. It follows that students are becoming more discriminating consumers. In making choices between universities, they are placing greater weight on securing valuable future employment. In response, institutions are competing to provide more relevant knowledge and skills through more effective teaching and learning.

Sweeping developments in the higher education market are now intensifying this competition. A global market has emerged. Many students are going abroad to study. Others look to the new, internationally available, digital platforms to provide or supplement their learning.

Taken together, these developments create a powerful demand for data to measure the quality of teaching and learning in higher education institutions around the world. Institutions need data to build on their competitive strengths and address weaknesses. Governments need data to determine policy and funding priorities. Employers need data to assess the value of qualifications from different institutions. Perhaps most importantly, students themselves need data. This is in part to help them make informed decisions about their preferred place of study. But it is also because in an age of widespread youth unemployment, students will value the opportunity to show prospective employers evidence of the levels they have attained in international assessments.

However, in a period when governance, accountability and transparency in higher education have been strengthened in many countries, and when data on the performance of institutions in research is acknowledged to be powerful and comprehensive, the loud demands for data on learning outcomes, expressed by students, institutional leaders, ministers and business, continue to go unmet. In this key area, there is a continuing and damaging absence of information of a quality to ground credible benchmarking and comparison.

Without data on learning outcomes, judgements about the quality of teaching and learning at higher education institutions will continue to be made on the basis of flawed rankings, derived not from outcomes, nor even outputs – but from idiosyncratic inputs and reputation surveys.

I believe that comparative measures of higher education learning outcomes hold significant promise. They allow *governments* to evaluate the quality of their university-educated human capital among the higher-educated cohorts against international standards. They enable *institutions* to compare and benchmark the learning outcomes of their students against international standards in order to improve the quality of teaching and learning. And they empower *students* to weigh their learned skills against the distribution of learning outcomes in their own institution and country and against international standards.

In a globalising world, governments want to have more profound knowledge about the education and skills pool at the upper end of the distribution. Economic arguments

relating to productivity, innovation, competitiveness and growth, and social arguments relating to social cohesion, trust and various other social outcomes of education, create a need for governments to assess the learning outcomes of their new cohorts of university-educated graduates. Furthermore, governments want to have the evidence that allows them to assess the effectiveness of public investment in university education.

Measures of learning outcomes will provide universities with profound insights into the effectiveness of teaching and learning and thus constitute a highly significant advance in the quality assurance environment. Each institution can obtain detailed data and analysis which highlight the learning outcomes of their students against a range of national and international benchmarks.

Last but not least, individual student feedback on learning outcomes can be critical to institutional and student engagement. By obtaining internationally-comparative data on learning outcomes, students obtain a way to ascertain their knowledge and skills independently. Students, teaching staff and institutions can also see how the performance of their institution compares with that of other institutions, nationally, internationally and against institutions with similar characteristics.

How?

The 'how' question is a lot tougher. What skills should we value, measure and compare? Universities are discovering their own

answers to this, as every university has their own profile. But labour demand is relevant too, and the things that are easy to teach and easy to test are also easy to digitise, automate and outsource.

One obvious answer to this is to assess learning outcomes in academic disciplines, and numerous approaches exist that demonstrate this is feasible. Any framework for the measurement of learning outcomes would therefore include learning outcomes in disciplinary contexts. These are easily interpretable in the context of departments and faculties. But there are challenges too. Such measures require highly differentiated instruments which reflect international agreement, and they are likely to exclude disciplinary areas that are not amenable to large-scale assessment or not sufficiently invariant across cultures and languages.

There is no question that state-of-the-art knowledge in a discipline will always remain important. Innovative or creative people generally have specialised skills in a field of knowledge or a practice. But transversal skills are equally important, such as ways of thinking, involving creativity, critical thinking, problem-solving and decision-making; ways of working, including communication and collaboration; tools for working, including the capacity to recognise and exploit the potential of new technologies; and the social and emotional skills that help people live and work together.

Conventionally, our approach to problems was breaking them down into manageable bits and pieces and then teaching students the techniques to solve them. But today we create

value by synthesising the disparate bits. This is about curiosity, open-mindedness, making connections between ideas that previously seemed unrelated, which requires being familiar with and receptive to knowledge in fields other than our own. If we spend our whole life in the silo of a single discipline, we will not gain the imaginative skills to connect the dots for where the next invention will come from.

Perhaps most importantly, students still learn mostly individually and at the end of the term, we certify their individual achievements. But the more interdependent the world becomes, the more we rely on great collaborators and orchestrators who are able to join others in life, work and citizenship. Innovation, too, is now rarely the product of individuals working in isolation but an outcome of how we mobilise, share and link knowledge. So universities need to prepare students for a world in which many people need to collaborate with people of diverse cultural origins, and appreciate different ideas, perspectives and values; a world in which people need to decide how to trust and collaborate across such differences; and a world in which their lives will be affected by issues that transcend national boundaries.

An assessment of learning outcomes that goes beyond disciplines and also includes transversal skills would be more attractive for employers. It would also transcend disciplines and institutions, and therefore allow for a much wider range of comparisons. Not least, it could become a powerful driver for improving the quality of teaching in different disciplines. But there are challenges too. First of all, transversal skills tend to reflect cumulative learning outcomes and need to

be related to prior learning. They also do not relate to the kind of subject-matter competencies that many universities, departments or faculties would consider their province. There are also conflicting views on the assessment of generic skills. Some view generic skills as inseparable from domain. Many such skills are common or transversal to several domains of study, but express themselves through the domain-specific environment. For example, critical thinking is a transversal skill, but can only be acquired through the deep mastery of a given field of knowledge, from which associations can be made with other domains. Thus economists and civil engineers, for example, both use communication and problem-solving skills, but in different ways. In the Feasibility Study, both types of skill were assessed in the economics and civil engineering strands and the Main Study would take the same approach, in these and other domains. Others view generic skills as separate from domain, seeing the skills of communication and problem solving as discrete and applying to students in all domains.

However, if we combine disciplinary measures with transversal ones, we can obtain a solid basis for comparing learning outcomes.

The next question we run into is: are we going to talk about systems or institutions? Governments generally want to know how their higher education systems are performing. But there are major obstacles to this: institutional missions – and their student intakes – are highly varied. There is considerable variation in institutional structures across countries. Nationally representative samples seem unrealistic as long as one sticks to the voluntary nature of such comparisons, which I consider

essential. Mandated assessment, even if it were possible, would not be effective as a tool for improvement at the level of service provision. Last but not least, large cross-country differences in enrolment rates raise questions about the interpretation of comparative performance measures. Getting at system-level assessments seems, in the near-term at least, an impossible task. The more relevant alternative is to focus on institutions. Moreover, universities increasingly see themselves as part of a global landscape. They talk to peers in other countries and are not necessarily mainly interested in how they compare with neighbouring colleges. That lends itself to using the institution as the unit of comparison and analysis, rather than speaking of higher education systems. It is not so easy for governments to get used to the idea that the user decides and that these are voluntary arrangements, but that is probably the only way I can see this working in the foreseeable future.

There is a third challenge too: what is the nature of the metric against which we want to compare institutional performance? Individuals, whether prospective students or employers, would likely want to know the 'bottom line' of the performance of institutions, departments or faculties, irrespective of the conditions under which these were obtained. In contrast, institutions and policy makers wishing to assess the quality of services provided would be mainly interested in the 'value added' by the institutions. Any promising approach to measuring learning outcomes would need to combine both perspectives. Logistically and methodologically, measuring value added through longitudinal metrics is very difficult. But it is relatively easy to provide analytical value-added measures based on relating statistically individual students' learning

outcomes to socio-economic and institutional background variables. So, to make comparisons as meaningful as possible, detailed contextual information can be collected from students and institutions. This means that it will be possible to compare the performance of students against those with similar characteristics.

Criteria for success

That brings me to the third question: what are the criteria for success to evaluate the relevance and validity of the outcome measures? I would expect several things from international measures of higher education learning outcomes.

First, they should reflect central and enduring parts of higher education teaching that relate to quality of outcomes. They should reflect aspects that can be improved, and they should be appropriate across cultures and valid across institutions and systems.

A more difficult question is how to balance the breadth and depth of any metrics. Clearly, some focus is important but one would need to avoid tunnel vision and give educators the depth needed to stimulate improvement. Take the school system in England as an example: to an outsider, it often looks like a mile wide but an inch deep. There are a lot of things being taught but actually not at the level of cognitive demand, rigour, focus and coherence that you would see in some parts of East Asia. They tend to teach fewer things but at great depth and with a high level of cognitive demand.

It is also important to find a good balance between outcomes and process. The design and implementation of an assessment is important, but so is the process of communication with key stakeholders on the nature and value of the assessment and the information gains. Giving faculty meaningful feedback on the quality of teaching should be a central objective of any assessment of higher education learning outcomes.

Last but not least, a challenge will be to obtain measures that are as comparable as possible but as specific for institutions as necessary to be meaningful. One of the aspects of the green paper that I like is that institutions, which may not favour the comparative perspective, will have the opportunity to provide an institutional perspective too, giving a balance between the two.

What should we expect from the measurement process itself? Accountability is nearly always an insufficient reason for measuring learning outcomes. The key is that measurement supports improvement of learning at all levels of the university system. Measurement should also be largely performance based. Advances in assessment methodology are very promising here. In the most recent edition of PISA, we used digital assessment tools. Looking at whether students get the answer right or wrong on the test is interesting, but what makes it much more interesting is tracking how students get to the answers. We can follow the thinking processes of people as they respond to the test and, with the new assessment on collaborative skills and social skills, we can even track how they interact with other people. That is what creates pedagogical value. Measurement today can make students' thinking visible and allow for divergent thinking.

Measurement should also be part of a comprehensive and well-aligned continuum, communicate what is expected, and it should add value for teaching and learning by providing information that can be acted on by students, teachers, and administrators.

I firmly believe that measurement should be adaptable and responsive to new developments. Some people say that you cannot measure change if you change the measure. But if you follow that logic, you freeze everything and you end up measuring progress on things that are no longer relevant. The important tension comes in the need to balance comparability over time against the desire for metrics that are relevant to our situation today.

The toughest question relates to validity, how do we know that our measures reflect what we actually value rather than what is just easily measurable? But there are several things we can do. We can look, for example, at how the results compare with direct and indirect measures on research outcomes. It is an interesting question to what extent those are correlated or not. Or we can look at institutional factors and non-cognitive characteristics that are known to be tied to successful study and achievement.

Alumni ratings are an easy win. Today, we go on eBay or Amazon and buy things from complete strangers, or we may rent out our apartments to complete strangers, because we have reputational metrics about the people from whom we buy. Looking at the alumni perspective could be another validity check. There are really interesting experiments around

that, including in your green paper, but I am sceptical that they are sufficient as a proxy for learning outcomes. I think it is great as a sort of variable to which you can correlate outcomes. But I do not think you can ever substitute learning outcomes with what people think about those learning outcomes.

We can also look at how university learning outcome measures relate to the labour market and social outcomes of higher education. Employment outcomes are really powerful for most countries. You can see the risk of unemployment is a lot higher for people without great skills and a lot lower for people with university qualifications. If you look at the newspapers in Spain, for example, they suggest university education is worthless. Actually, while there are a lot of unemployed university graduates in Spain, your risk of unemployment is a lot higher if you do not have a degree. The picture is similar on earnings. Look at men and women gaining tertiary qualifications and how much they earn. Again, such indicators are not a substitute for learning outcomes because you do not know to what extent the picture is due to the demand for skills or the supply of skills. But it is interesting to see that, after discounting all expenditure including tuition fees, there is a lot of money left for university graduates. It is between $200,000 and $400,000 (USD) on average in some countries – much more than they spend on higher education. It is positive in every country.

Despite the rapid decrease in knowledge workers in virtually every OECD country, this has not led to a decline in the labour market value of qualifications. That tells us that the demand for better skills is rising faster than the supply. At least so far. Nobody knows how long this will continue. But maybe

if we had asked ourselves the same question 100 years ago for school education, we would have had the same debate. How many people do we really want to complete school? How many of them do we really need? Maybe we are living in a time where tertiary education, whether it is university or other forms of tertiary education, is becoming the norm. Also taxpayers benefit. You can see that in most countries, taxpayers take between $100,000 and $200,000 dollars more out of every university graduate than what they invest. It is a very good reason for government to stay actively involved in the financing of higher education.

There are also some major methodological challenges beyond those mentioned so far. For example:

- Can we drink from the firehose of increasing data streams that arise from new assessment modes?
- What is the right mix of crowd wisdom and traditional validity information?
- Can we sufficiently distinguish the role of context from that of the underlying cognitive construct?
- Can we utilise new technologies to gain more information from students without overwhelming students with more assessments?
- How can we create assessments that are activators of students' own learning?

These difficulties are real but I am confident that we can address most of them with available methodologies. I think we

can build measures of learning outcomes that meet the tests of: coherence by building on a well-structured conceptual as the foundation for assessments; comprehensiveness, in terms of using a range of assessment methods to ensure adequate measurement of intended constructs and measures of different granularity to serve different decision-making needs and in terms of providing productive feedback, at appropriate levels of detail, to fuel accountability and improve decisions at multiple levels; and continuity, in terms of providing a continuous stream of relevant evidence.

Conclusion

Let me conclude by stating that student learning outcomes should be in the critical path of assessing the outcomes of higher education. I really do not believe that we can find any shortcut to measuring the quality of higher education that bypasses students and student learning outcomes. We can find proxies and variables that correlate, but at the end of the day it is for learning gains that we go to university. Let me also affirm that we can measure a sufficient range of learning outcomes in appropriate, valid and reliable ways to make such efforts worthwhile.

Of course, the political economy of all this is tough. Whenever we look into the mirror of what others show can be achieved, we may not look as beautiful as we thought, or as beautiful as others have told us we are. So it is unsurprising that the biggest opposition comes from those countries and institutions that fear they have most to lose because they fear their outcomes

may not be as good as their reputation suggests. Those countries and institutions have loud voices but it just goes to show we should be trying harder.

Trustees

Professor Sir Ivor Crewe (Chair)
Dame Sandra Burslem
Professor Sir Peter Scott
Dr Ruth Thompson
Professor Sir Nigel Thrift

Advisory Board

Professor Janet Beer
Sir David Bell
Professor Sir David Eastwood
Professor Dame Julia Goodfellow
Professor David Maguire
Professor Dame Helen Wallace

Partners

BPP University
Ellucian
Elsevier
HEFCE
Higher Education Academy
Jisc
Kaplan
Mills & Reeve LLP
Pearson
Times Higher Education
UPP Group Limited
Wiley

President

Bahram Bekhradnia

The quality of teaching and learning in higher education is under enormous scrutiny. For example, in England there is to be a new Teaching Excellence Framework.

In this revised version of the 12th HEPI Annual Lecture, which was originally delivered in December 2015, Andreas Schleicher of the OECD makes a powerful case for directly measuring the learning gain of students on a comparable basis across the world for the first time.

While recognising the technical challenges involved, the lecture argues this is now possible and, because the proxies for learning gain have such big problems, there is no substitute for starting to measure it directly.

HEPI was established in 2002 to influence the higher education debate with evidence.
We are UK-wide, independent and non-partisan.
January 2016 ISBN: 978-1-908240-11-8

Higher Education Policy Institute
99 Banbury Road, Oxford OX2 6JX
Tel: 01865 284450 www.hepi.ac.uk

Printed in the UK by Oxuniprint, Oxford
Typesetting: Steve Billington, www.jarmanassociates.co.uk