Evidence-based policy and higher education quality assurance: progress, pitfalls and promise

Beerkens, M.

Evidence-based policy and higher education quality assurance: progress, pitfalls and promise

Maarja Beerkens

Published online: 21 May 2018.

Submit your article to this journal ↗

Article views: 1273

View Crossmark data ↗

Routledge
Taylor & Francis Group

Evidence-based policy and higher education quality assurance: progress, pitfalls and promise

Maarja Beerkens

Institute of Public Administration, Leiden University, Den Haag, Netherlands

ABSTRACT

Evidence-based policy has become a norm in the current policy-making rhetoric, affecting also higher education quality assurance. This article agrees with critics that rigorous ex-post impact studies are highly challenging in the field of quality assurance. Nevertheless, there are alternative ways how evidence can effectively guide quality assurance policies and how evidence-based mentality can be encouraged by government policies. A more realistic view on how evidence informs policies (indirectly and via stakeholders' arguments) and how professionals incorporate evidence in their work (selectively and next to other information sources) broadens the scope for useful evidence for higher education quality assurance.

Introduction

'Impact' of external quality assurance in higher education has received much attention in recent years, both in practice and in academic literature. The key professional associations have started to take the issue seriously. The *European Association for Quality Assurance in Higher Education* (ENQA) set up a multi-year working group to explore how to measure impact (ENQA 2016) and the *International Network for Quality Assurance Agencies in Higher Education* (INQAAHE) in their last annual meeting discussed the question of what works in quality assurance, and how do we know it? The interest in the question of 'impact' is reflected also in academic literature. *Google Scholar* indicates a rapid growth of articles with keywords 'impact' and 'quality assurance' in higher education literature, from 7 articles published in 2000, to 38 in 2010 and 84 in 2015.

Reasons for the growing interest in 'impact' and 'evidence' in quality assurance are probably manifold. First, the field of quality assurance has evolved significantly over the last two decades. As Stensaker (2007) argues, the era of enthusiasm is replaced by the era of realism in higher education quality assurance. As quality assurance is no longer a novel practice encountered with eagerness, there is time and a need to reflect what impact past activities have had. Occasional signs of fatigue in the universities facing a

CONTACT Maarja Beerkens ✉ m.beerkens@fgga.leidenuniv.nl 🖂 Institute of Public Administration, Leiden University, Wijnhaven Building, Turfmarkt 99, Den Haag 2511, Netherlands

new evaluation round (Westerheijden 2007) demand a 'proof' of the positive effect that quality assurance creates for higher education. Secondly, the field of quality assurance is professionalizing. Knowledge and research about quality assurance have evolved considerably over time, training of evaluators has become a common practice, and quality assurance agencies have been not only accumulating knowledge but are actively sharing their expertise internationally. As the level of sophistication rises and quality assurance is increasingly perceived as a profession requiring technical expertise, there is also a need for hard evidence on what works and what does not work. Thirdly, growing interest in evidence is universal in the policy sphere. Evidence-based policy-making is a concept with a strong appeal, rapidly spreading beyond its origins in health and social work. Also, higher education quality assurance feels the pressure of the paradigm.

Despite of all the interest in impact studies of higher education quality assurance, the state of the art in the field is often found lacking. Stensaker (2007) claims the field to be in its infancy and Harvey (2016) looking back at the twenty years of the journal *Quality in Higher Education* names 'the failure to adequately explore impact of quality assurance' as one of the key conclusions. To be sure, there is no lack of 'evidence' out there. Quality agencies evaluate their activities regularly. Many reports by these agencies and other organizations analyze the state of the higher education sector in general, various surveys on stakeholder satisfaction, graduate employability and graduate satisfaction exist, and studies on other specific aspects of higher education are common (Damian, Grifoll, and Rigbers 2015). But what impact various quality assurance policies have on student learning – which is presumably the primary target of the quality policies – is to a large extent unknown.

This article explores the notion of 'evidence-based policy' in the context of higher education quality assurance. We examine various ways how an external quality assurance system can use an evidence-based approach in their activities, and what kind of evidence is needed to support the approach. Extensive experience with evidence-based policy-making in various public services over the last two decades has curbed the simplistic hope that rigorously designed impact studies on policy instruments give the final answer for better policies. The phrase 'evidence-based' policy-making is increasingly replaced by 'evidence-informed' policy-making, realizing the importance of practical (professional) and political knowledge next to scientific knowledge for making policy decisions. Experience with evidence-based policy-making thus calls for a broader look at what constitutes evidence and how evidence contributes to the policy process and to quality improvement more broadly. What this broader look means for higher education research is discussed in the second part of the article.

The effects of external quality assurance: what do we know?

In about two decades, external quality assurance has developed to a well-institutionalized regulatory field (Westerheijden, Stensaker, and Rosa 2007). Quality assurance agencies have been established virtually all over the world, conducting accreditation, assessment and audit activities in various forms and regularity. It is fair to ask what this quality movement has contributed to higher education. Effects probably vary across different countries and depend on a specific quality assurance instrument but accumulating evidence primarily from Europe refers to some commonly reported outcomes (e.g. Brennan and Shah

2000; Harvey and Williams 2010; Liu, Tan, and Meng 2015; Minelli et al. 2006; Rosa, Tavares, and Amaral 2006; Stensaker 2007, 2003).

It has been well established that external quality assurance seems to have affected internal governance structures in universities. It has strengthened the position of central administration within universities and it has contributed to managerial power. Quality control at the programme level necessarily creates new accountability relations between the central administration and lower level units responsible for delivering education. Furthermore, external quality assurance has contributed to strategic management within universities and to creating a common identity.

Perhaps most noticeably, external quality assurance has helped to professionalize quality processes within universities. This means developing an appropriate bureaucracy in the form of standardized processes and norms and by creating special organizational units or positions to facilitate the tasks. To be fair, quality assurance has created also procedures, demands and organizational structures that are perceived as burdensome and ineffective by those in lower levels of the organization. As part of the professionalization, accumulating knowledge has contributed to organizational learning and sharing of best practices both within and between institutions.

The main question is if all this – stronger leadership, management, new units and formalized procedures – has made education any better. Evidence in some countries shows that external quality assurance has helped to establish new formal standards (e.g. modularization of education in Germany), although the extent to which this created a qualitative change may be debatable (Suchanek et al. 2012). More broadly, external quality assurance has helped putting educational quality on the agenda, thereby increasing its importance and somewhat balancing the dominance of research performance as the key focus within universities. In other words, it may have been a trigger for a change in attitudes, a cultural change. It seems to have contributed to collegial discussions about curriculum (e.g. when developing a self-study report or as a preparation for an on-site visit). Thereby we might expect that it has contributed to coherence of programmes and stimulated reflective practice. Performance data also seems to show improvement as a response to quality assurance. Retention rates, graduation rates, the level of final awards (in the UK system), and graduate employment seems to have improved in the systems where the performance data was subject to evaluation, showing that universities are indeed responsive to external incentives. Banta (2010) zooms closer on educational practices in the US and observes improvements in pedagogical practices, student advising and learning communities.

There is thus quite a lot of evidence on positive (and some negative) effects of external quality assurance on universities but the question of whether graduates now walk out with better knowledge and skills as a result of all the quality reforms is still hard to counter. Clear evidence on the impact of quality reforms on student learning would be certainly helpful to overcome occasional criticism and to develop effective quality policies for the future. Establishing such evidence, however, is very difficult on several reasons. There are a few challenges that such impact studies would face, most importantly related to questionable causality, heterogeneous effects, and imperfect measurement.

First, universities are complex organizations and it is often impossible to determine what exactly has caused changes. In the case of effects reported above, an external quality assurance instrument has often functioned as a 'catalyzer' for change, an external impetus to legitimize cultural and organizational reforms internally (Harvey 2006). Dill

(2010) argues that the most important contribution of quality audits was to aid senior managers to initiate quality assurance systems within their institution and to provide system-wide information on best practices. An organizational change in universities is an iterative process, not a simple linear trajectory from one single cause to one single effect (Stensaker 2003, 2008). Reality gives rarely an opportunity to study empirically what would the current situation be without being subject to external quality assurance (i.e. quasi-experiments with a control group are rarely available).

Furthermore, observed effects may not be constant over time. It is well reported that quality assurance instruments have had a major effect in the first round of the exercise – it has helped to raise awareness, to change 'quality culture', and to initiate discussions that were not present before. Such positive effects can be much smaller in later years. On the other hand, a limited effect of the next assessment round does not mean that the system would be just as well without any external quality assurance. An effective quality assurance system may be doing its work quietly behind the scene, providing sufficient stimulus for universities and programmes to take quality seriously and not letting the system deteriorate. Moreover, the effects can be influenced by other contextual factors – other incentives present in the system, alignment with a university's mission and goals – which makes generalization of the results difficult not only over time and between countries but also between institutions and organizational units.

Most importantly, though, the problem lies in measuring student learning. The gold standard in evidence-based policy is an explicit measurement of the effect of a policy intervention on final outcomes, preferably by randomized trials to minimize the effect of potential competing causes, self-selection and other biases. Such an approach in higher education is however very difficult. Namely, it is difficult to measure the outcome – educational quality – in the first place, even more so to measure change in educational quality and the role of quality assurance in it. Skeptical voices of higher education quality assurance raise namely this concern: all the accumulating evidence demonstrates that procedures and processes have changed, awareness about and attention to quality has changed, and perhaps even pedagogical practices have changed but are our graduates now better educated as a result? A significant part of educational research of course measures learning outcomes to evaluate the effects of various pedagogical approaches and instructional techniques. But such studies are usually limited to very specific outcomes and usually within one course. To estimate the impact of quality assurance mechanisms, reliable data is needed for overall learning during a university programme and over time.

Some ambitious initiatives attempt to develop instruments for measuring student learning more comprehensively. Collegiate Learning Assessment (CLA) in the US is one such instrument that assesses some generic skills such as written communication, critical thinking, analytic reasoning and problem solving, and it is developed with a purpose to measure institutional performance and particularly the value a university adds to students' development (Klein et al. 2007). The famous AHELO project of the OECD was an explicit attempt to measure learning outcomes, and the value added by universities to students' knowledge and competences, in an internationally comparable way. Existence of such data would certainly stimulate studies on the effects of various policy instruments and organizational measures on learning quality. However, as the problematic and controversial results of the AHELO project show, it is not only politically and financially difficult but also methodologically challenging to create a reliable and valid assessment instrument for

higher education learning outcomes (OECD 2013). The experiences also raise doubts of whether seeking such an ideal is at all a realistic goal and whether the benefits exceed the high costs of such activities.

While rigorous studies on impact of various quality 'interventions' are rare and highly challenging, there is no reason to be skeptical about the role of evidence in enhancing higher education quality. There are different ways how an external quality assurance system can rely on an evidence-based approach in their activities. Learning from impact studies – the gold standard of evidence-based policy – is only one way and not necessarily the most effective way in this specific case. In the next section, we will discuss different ways how the principles of evidence-based decision making have been incorporated in external quality assurance.

Evidence-based approaches for external quality assurance

Evidence about the link between activities undertaken and their effect on learning quality is essential for evidence-based quality enhancement for higher education. What kind of activities are under scrutiny for creating evidence and how this link is incorporated into an external quality assurance system may take different forms though. Below we will discuss four models how an evidence-based approach is integrated into quality assurance activities.

Impact studies on quality assurance instruments

As discussed above, systematic impact studies on quality assurance instruments are rare but there are nevertheless examples how such an approach is used for policy development. Probably the best example comes from a neighboring area, research assessment. The UK research assessment has used various performance indicators to monitor universities' research output. Numerous empirical studies have evaluated the effects of the assessment on research output and on research behavior, and the choice of indicators has been continuously adjusted to minimize distortive effects (Barker 2007). Granted, policy adjustments are never based solely on conclusions of an impact study but on broad-based consultation with various stakeholders. Nevertheless, research evidence plays an important role by being incorporated in views and arguments of various stakeholders (Barker 2007).

Examples of monitoring the effects of higher education policy instruments on universities' performance come primarily from the UK and Australia. Not coincidentally, these quality assurance mechanisms are based on performance metrics, which makes observing effects relatively easy as data collection is part of the instrument itself. In Australia, the early years of using performance indicators showed that universities indeed responded to the instrument and performance on several dimensions (e.g. retention rate, employability) seemed to improve. The question remains if this improvement indicates a qualitative change in education or universities just address the indicators as a goal of its own. As accumulating research demonstrated technical difficulties of performance indicators, the Australian government undertook and commissioned a variety of benchmarking and performance indicator studies to evaluate the sector across a broad range of activities (Neumann and Guthrie 2009).

Some accreditation schemes have been also subject to impact evaluation. CACREP (Council for Accreditation of Counseling & Related Educational Programs), a professional association that accredits programmes in counseling and related educational fields in the US, announces on their website confidently, 'research shows CACREP graduates perform better on the National Counselor Examination for Licensure and Certification' (CACREP 2017). Indeed, studies show not only that accreditation is *perceived* by teachers and coordinators to have positive effects on curriculum and educational quality (Holcomb, Bryan, and Rahill 2002), but also that graduates of CACREP accredited programmes are less likely to be sanctioned for ethical misconduct, which is one of the key focus areas of the accreditation (Even and Robinson 2013), and perform better at the national exam (Adams 2006). Nevertheless, it is difficult to determine the causal effect of the accreditation procedure, apart from self-selection of students to accredited programmes, or self-selection of programmes applying for the accreditation.

Main technical and feasibility challenges of such impact studies were discussed above and we turn now to another view how evidence can be incorporated in a policy.

Evidence-based quality assurance instruments

Measuring the true effect of a quality instrument on student achievement is nearly impossible or at least highly resource-intensive. A more effective and practical approach could be an attempt to design a quality assurance instrument that is strongly based on an existing evidence base. Relevant evidence is thus collected not after the fact to test its working, but ex-ante evidence can feed directly into a quality assurance instrument. To a certain extent hopefully all quality assurance instruments rely on some evidence on what factors matter for educational quality. Often, however, quality instruments are built on an intuitive *rational* approach of what should matter for quality rather than on an *empirical* approach of what is proven to matter (Pascarella 2001).

Some quality assurance instruments indeed take such an empirical approach seriously. The National Survey of Student Engagement (NSSE) in the US is a good example. This is a carefully designed instrument that relies on student-reported data on their learning experience, and it is developed based on existing empirical evidence on what activities and practices contribute to student learning (Ewell 2010). The instrument is developed around a convincing set of empirical evidence that student engagement and effective learning environment are the key contributors to student achievement and development. Drawing from this evidence, the five NSSE benchmarks that serve as a framework around which the annual reports are organized are: level of academic challenge, active and collaborative learning, student–faculty interaction, enriching educational experiences, and supportive campus environment. An emphasis on student engagement in external quality assurance as a proven mechanism to student learning is gaining popularity. Many researchers argue that quality assurance needs to take account of student engagement and learning (Coates 2005; Meyer 1999). The centrality of student-centered learning in the *Standards and Guidelines for Quality Assurance in the European Higher Education Area (ESG)* (ENQA 2015) could also be an example of such an approach, although it is unclear to what extent the focus is really based on evidence as opposed to ideological stand, and whether its formulation is specific enough to deliver expected benefits.

Such an engagement-based approach addresses one of the critical issues in quality assurance. Many quality instruments focus on performance indicators or process reviews, but they rarely link to student learning directly (Harvey 2016). The *Teaching and Learning Quality Process Review* in Hong Kong is often praised as one of the few instruments that puts a strong focus on pedagogy (Massy and French 2001). Such an approach, however, does not come without its own challenges. Technical feasibility is one of them. While NSSE is an exemplary initiative in its rigour and conceptual development, ex-post studies on the relationship between student engagement as recorded in the survey and student performance (retention, average grade) exhibit actually little correlation. Carini, Kuh, and Klein (2006) show a link between engagement measures and educational outcomes but the effect sizes are very small, although somewhat higher for the lowest ability students. Similarly, Gordon, Ludlum, and Hoey (2008) observe that student responses on the NSSE do not predict their success in terms of their grade or retention. Getting to relevant and reliable predictors of success seems to remain a challenge.

Furthermore, such a deterministic view that one type of 'technology' leads to most effective learning may be somewhat simplified. Bramming (2007) makes a convincing case for transformative learning and calls for 'philosophically grounded pedagogies' to support the goal. Her argument illustrates that pedagogies or educational approaches are not value-free, simply more or less effective in objective terms. They are grounded in a certain view about the student, the educational ideal and often carried by the 'Zeitgeist'. This may make the next approach of using evidence for quality approach more appealing for effective quality assurance.

Encouraging evidence-based approach within universities

Some quality assurance instruments take a position that programmes should demonstrate an evidence-based approach in their educational planning and programme development. The programmes themselves have the responsibility to monitor student learning and observe the achievement of standards. Programmes should also use this evidence for making adjustments and improving the quality of education. What is being assessed thus is not only achieving the standards but an effective evidence-based approach for internal management.

A good example here is the audit by the Teacher Education Accreditation Council (TEAC), an instrument that focuses on sound evidence generated and used by the programme. The accreditation is based on explicit standards and detailed evidence that standards are met (El-Khawas 2010). Evidence must be produced on student learning, on valid assessment of student learning, and on the use of evidence for academic planning and continuous improvement of educational quality of the programme. The programmes, however, have flexibility in what evidence to provide on student learning, e.g. ratings of portfolios of academic accomplishment, performance of pupils taught by graduates, graduates' self-assessment, employers' evaluations, etc. This approach allows programmes to gather detailed evidence that is both meaningful and useful for improvement in a specific unit (El-Khawas 2010). A programme is also obliged to think critically whether the evidence is dependable, persuasive and representative of the programme, and the review panel has the liberty to interpret the evidence and perhaps come to alternative

conclusions. According to El-Khawas (2010), the initial accreditation round was for some programmes quite challenging since they needed to undertake studies as part of evidence collection. However, participants generally claim that the accreditation process indeed fostered improvement, as opposed to being done as an administrative task to obtain the certificate.

In short, a quality assurance process can encourage university programmes to take an evidence-based approach. Instead of checking and evaluating whether a programme meets certain standards as a snapshot, the emphasis is on programmes themselves taking a proactive approach to monitoring student learning and demonstrating an evidence-based approach in academic planning.

Experimental approach to curriculum and course design

An evidence-based approach within a programme can go also a level lower and concern specific pedagogical practices and teaching modes. It is an experimental approach to curriculum development whereby the effects of new teaching methods or changes in curriculum design are systematically evaluated. Here the purpose of evidence is to get some light into the black box of student learning and examine what response certain pedagogical choices induce. This may for example concern the effects of class size on learning, or the number of contact hours (i.e. time spent on interaction between teacher and students) on learning. A lot of educational research exists on effects of various didactic practices and techniques. However, often evidence is inconclusive, conditional and context specific (e.g. evidence on class size) and therefore it may be more fruitful to examine effects within a specific programme for a specific course.

Evidence on the effectiveness of certain pedagogical techniques and teaching approaches can effectively inform improvements in a programme. Innovation using information technology, for example, is one area where evidence is important to avoid either going along with a hype or rejecting good ideas due to prejudice and unfamiliarity. ICT in teaching may encounter some resistance from academic staff who themselves have not been exposed to such teaching practices, but evidence can convince not only about the effectiveness of such practices for student learning but about cost-efficiency that is highly needed under increasing resource constraints in higher education. As many programmes are experimenting with alternative teaching modes, such as flipped classroom or online learning environments for a specific course, it is possible to study the effects of such changes in a specific context. Bowen et al. (2014) is an example of a rigorous and large-scale study evaluating the outcomes of switching an introductory statistics course to a partly online format. The study indicates no significant difference in learning outcomes while producing significant savings in terms of instructional costs. Such evidence is thus valuable for designing an effective higher education programme and to use scarce resources with maximum efficiency.

This model sets quite high demands on individual staff members to take an evidence-based approach to their own teaching tasks but there are reasons to be optimistic about increasing 'professionalization' of teaching. In several countries teachers are now required to take some training and obtain a certificate to demonstrate their teaching competence. University-wide teaching and learning centers advise faculty members in advancing their

teaching practices, and PhD-training increasingly involves preparation for the teaching profession. It can therefore be expected that the use of evidence about effective teaching practices will be more common among individual teachers.

Educational research of course offers much evidence about the effectiveness of various pedagogical approaches, assessment methods and technology use. The overwhelming quantity of evidence, however, is also the problem for using this evidence effectively. Many are critical about poor evidence-base for policy decisions, but the (poor) use of evidence by teachers illustrates nicely the difficulty of bridging the gap from scholarly evidence to daily practice. The last two decades of experience with evidence-based policy-making in various policy sectors has pushed the field for a broader and more realistic look to evidence. Instead of a primary focus on rigorously conducted impact measurement, the questions like what constitutes evidence, how the evidence is used in policy making, and who should be involved in evidence-based policy making have become more important. In the next section, we will see what the evidence-based policy movement has learned about the role of evidence in actual policy making, and what lessons higher education quality assurance can draw from the experience.

'Evidence-based' vs 'evidence-informed' policy

'Evidence-based policy' as a policy paradigm has enjoyed a major success from the late 1990s onward. The 'What-works-is-what-matters' attitude is appealing, and it fits the era when effectiveness and efficiency goals are at the forefront and ideological disagreements are considered more as a nuisance. Nevertheless, only in very exceptional cases is the relationship between evidence and policy-making a simple, linear process where research results directly lead to policy changes. John Maynard Keynes has been claimed to say, ' … there is nothing a government hates more than to be well-informed; for it makes the process of arriving at decision much more complicated and difficult' (from Skidelsky 1992, 630). It may sound sarcastic, but it also illustrates the fact that evidence alone rarely offers the final answer to complex policy problems in a political setting.

The contribution of evidence depends on the type of a policy issue (Head 2008). Policy issues vary considerably. There are simple issues that are clearly delineated, solvable with accurate background information and thereby technocratic in nature. In such cases evidence can indeed be easily absorbed in the decision-making and policy responses can be guided primarily by evidence. In a narrow sense, the effect of pedagogical practices on student learning can fall under this category, if we agree what student learning entails. In such cases, the main challenge for evidence-based policy making is an effective synthesis of many studies of varying scope and methodology, conducted in different contexts and probably reaching somewhat incompatible conclusions (Pascarella 2001; Young et al. 2002).

Many aspects of quality assurance, however, are significantly more complicated. They are cross-cutting and complex, often bound to trade-offs and normative choices. Studies on what works for student learning is invaluable for improving teaching and learning, but quality assurance tends to be linked to broader issues such as a variety of educational goals, differently perceived problems by the public, and interlinked policies (Beerkens 2015). In case of such complex issues, decisions cannot be derived from observing what works and what does not work. In such cases decision-making tends to be more relational and

negotiated, more accepting of different sources of knowledge (Young et al. 2002). Scientific evidence, particularly (quasi-)experimental and quantitative evidence, remains a 'gold standard' of evidence. Advantages of such knowledge are strongest for simple, technocratic issues. In a complex setting, scientific evidence takes a more interpretative character, and may require a form that is closer to real life complexities, such as action research. Furthermore, scientific evidence is just one type of evidence. Experiences within the evidence-based policy movement have made to question critically what counts as evidence and to recognize multiple evidence bases.

Head (2008) distinguishes between three sources of knowledge as relevant for evidence-based policy making: scientific knowledge, professional (practitioners') knowledge, and political knowledge. It is well accepted now that next to scientific results, practical wisdom of professionals of their specific communities of practice is important for interpreting scientific evidence and providing a necessary link towards effective policy or programme design. Similarly, the practical experience of professionals involved in quality assurance and faculty members involved in teaching and development tasks are an important source of knowledge. A professional source of knowledge needs to be recognized also in case of those delivering education in universities. The quote from Eric Ashby (1963) referring to a paradox that academics themselves do not take an evidence-based approach to their teaching tasks has become famous (see Brown 2013; Dill 1999):

> All over the country these groups of scholars, who would not make a decision about the shape of a leaf or the derivation of a word or the author of a manuscript without painstakingly assembling the evidence, make decisions about … staff-student ratios, content of courses, and similar issues, based on dubious assumptions, scrappy data, and mere hunch. (Ashby 1963, 93)

However, when it comes to scholars as practitioners of a teaching craft, they are not – and probably should not be – that different than masters of other professions.

> Ask any professional how they get to know what they need to know in their work and you get very diverse answers about the sources of their knowledge. Experience features large, both direct personal experience and that of colleagues. Most professions also have shared norms, values, ideas in good currency, sometimes articulated and made explicit, often tacit. Then there are the results of relevant research which may reach them in diverse ways. […] even with the most sophisticated hardware and software – the availability of all relevant knowledge is a hit and miss affair. (Solesbury 2001, 8)

In other words, academics cannot be expected to approach their teaching entirely from a scientific perspective.

Lastly, political knowledge is also important for successful implementation, for communication and ideological acceptability, and for defining the menu of acceptable alternatives. For example, learning outcomes is not self-evidently definable and measurable, and the political element in defining the purpose of education (e.g. labour market needs, civic responsibility, all-rounded development, etc.), and weighing tradeoffs between various goals (e.g. quality, access) is inevitable (Beerkens 2015).

Different sources of knowledge mean that evidence contributes to decision making not in isolation but *via* communication through policy networks and communities. Analyses of the evolution of performance indicators in research assessment in the UK (Barker 2007) or higher education policy plans in Australia (DETAG 2015), for

example, illustrate clearly how scientific evidence (selectively) is transformed into policies *via* extensive consultation rounds and political decision-making. Stakeholder engagement has become a norm in higher education quality assurance, and stakeholder engagement – somewhat counter-intuitively perhaps – can stimulate the use of evidence, through the process of creating a common understanding of problems and proposing changes (Beerkens and Udam 2017).

It is therefore important to recognize that the relationship between evidence and policy can take a variety of forms. Young et al. (2002), for example, distinguish between five, partly overlapping models for conceptualizing how (scientific) evidence and policy are linked. While a *knowledge driven model* assumes that research leads policy by pointing to and defining problems, the *problem-solving model turns* the relationship around and claims that research agendas are often determined by observing apparent and emerging policy issues. In an *interactive model* research and policy influence each other mutually *via* a policy community, while in a *tactical model* evidence is used by policy makers strategically and selectively to legitimate choices that have been already made as a result of political processes. And lastly, in the *enlightenment model,* research is at a distance, it helps to create a context for policy makers and a frame of thinking. All types of relationships are probably in play in higher education quality assurance, in different points of time.

Since evidence influences policy in different ways, there is a need for different types of studies and analyses. Impact studies are highly valuable for understanding the impact that a policy has on a system. However, 'what works' may not be the only question; but also how and under what conditions and circumstances an instrument works. As reminded also by Leiber, Stensaker, and Harvey (2015) there is a need for longitudinal and comparative studies addressing the impact of higher education quality assurance, not only to report effects but also to understand why and how the effects do or don't occur. This helps to anticipate heterogeneity of results in different contexts and evaluate alternative explanations for observed effects. Undeniably, countries can learn from each other's experience and potentially avoid some mistakes (e.g. Faber and Huisman 2003). One the other hand, there are several examples how quality assurance practices have been copied inappropriately from one country to another. This may be a case when a quality assurance system is copied from a large higher education system to a small system (Hopkin and Lee 2001; Hopkins 1990; Houston and Maniku 2005) or from a relatively wealthy system to a resource poor system (Ansah 2015). This also calls for theoretically grounded studies that draw from organizational and behavioral approaches to understand why an instrument works or not and under what assumptions (White 2009). Furthermore, as evidence contributes to policy making *via* policy networks and policy communities throughout the policy cycle, from agenda setting and policy formulation to decision-making and evaluation, effective evidence concerns not only impact but also problem definition. And not only evaluation studies are needed; also good descriptive, analytical, diagnostic, theoretical and prescriptive research is needed for evidence-informed policy-making (White 2009). Last but not least, the issue in practice is often not the lack of evidence but too much evidence with no professional consensus, which suggests a need for good synthesis studies that could inform policies. To avoid appearing too self-serving as a higher education researcher, this is not meant to suggest that all research is equally valid or relevant for informing policies or that

more and more research is needed, but only that useful evidence should not be defined too narrowly.

Conclusions

Evidence-based policy is a concept with a strong appeal, also in higher education quality assurance. Demand for impact studies in higher education quality assurance is not a surprise in a quickly professionalizing field that often faces criticism by powerful stakeholders. Sophisticated initiatives have emerged in recent years to measure student learning, suggesting that rigorous impact studies may become a more common practice in the nearest future. Nevertheless, as argued in this article, evidence-based approaches to quality assurance can take different forms. The 'gold standard' – rigorous (quasi-) experimental proof that a quality instrument has increased student learning – may be not only technically challenging and very costly but also not the most effective way to develop optimal quality assurance mechanisms and encourage quality education. 'It works!' is not the only evidence needed. How and under what conditions it works? What is the problem that needs addressing, if there is one? What are alternative solutions and would they work? All these are equally important questions that require an evidence-base.

There are other important ways how the evidence-based mentality can contribute to quality assurance. Next to impact studies, quality assurance bodies could also be more explicit in how their instruments are developed; what is the proof that what is being evaluated really matters for quality. But there is also another approach – to encourage evidence-based approach to teaching within universities. Universities can be encouraged to monitor critically their students' learning and collect evidence about their effectiveness in providing education. An evidence-based mentality could go down to the level of specific pedagogical practices or assessment methods within a single course.

It is important to distinguish between different purposes for collecting evidence. In an evidence-based policy approach, information is collected to understand effectiveness of quality assurance policies and to use the knowledge for adjusting and changing the design of the policy instrument. This rationale is different than collecting evidence for accountability purposes or to fix a 'market failure' in the system. There are concerns that universities may be getting complacent and that the quality of learning may be deteriorating in some cases, so universities are called to demonstrate their performance by measuring and monitoring student learning (Shavelson 2009). This links to fundamental information issues in higher education markets (Beerkens 2019). Since educational quality is not easily observable, incentives to provide quality education may be deteriorating in the system as a whole. External push to generate such information works as an accountability mechanism and can also restore incentives within the system. Collecting evidence in this framework of thinking is not *about* a policy instrument but *is itself* the policy instrument for improving quality in the system.

Regardless of whether we refer to (national) policies or course-level decision making, it is important to keep in mind that the process from a piece of evidence to a change in behavior or in policy is never simple and straightforward. Often the issue is not a lack of evidence but an overwhelming amount of diffused evidence. This is apparent also at the level of individual academics. As experienced in university level teaching and learning

support units, what is helpful for faculty members is not more knowledge on various techniques and pedagogical approaches, but an advice what to do in a specific course based on synthesized knowledge (Marincovich 2007). Like other professionals, academics develop their craft from various sources of knowledge from which scientific evidence is only one. Similarly, in the policy process evidence comes to life *via* perceptions, interpretations and arguments of various stakeholders involved in the process. Recognizing how evidence, in reality, is incorporated in daily practices and in a policy-making process may curb expectations but also broaden horizons about useful evidence.

Furthermore, organizational changes are a complex process. Approaching a quality assurance policy as a reductionist, linear 'intervention' that demonstrates specific measurable effects and that leads to choosing the best alternative, may be an oversimplification. Stensaker et al. (2011) come to a surprising conclusion in their study: the (perceived) effects of various quality instruments – evaluations, accreditation, audits – tend to be rather similar. This suggests that all the different policies could be seen as 'irritations' to the system that trigger a complex organizational change as a response.

Concluding with the wise words from an expert in evidence-based policy,

> evidence is an important part of the weaponry of those engaged in the discourse. … – to be effective – weapons must be handled with care. They must not be deployed casually or wastefully, and must always be used with as full regard to the risks for those who use them as to those against whom they are used. Knowledge is open to misuse quite as much as other sources of power. (Solesbury 2001)

Higher education quality assurance could probably make some very good use of this specific weapon, but with appropriate 'checks and balances', and under a 'civilian control'.

Acknowledgement

Disclosure statement

Funding

Notes on contributor

Maarja Beerkens is a Director of Studies and Assistant Professor in International Governance at Leiden University. Her research focuses on regulation and global policy issues. She has special

interest in higher education and science policy and in this area she publishes academically as well as conducts applied research for various national and international organisations.

References

Adams, Susan A. 2006. "Does CACREP Accreditation Make a Difference? A Look at NCE Results and Answers." *Journal of Professional Counseling, Practice, Theory, & Research* 34 (1/2): 60–76.

Ansah, Francis. 2015. "A Strategic Quality Assurance Framework in an African Higher Education Context." *Quality in Higher Education* 21 (2): 132–150.

Ashby, Eric. 1963. "Decision Making in the Academic World." In *Sociological Studies in British University Education*, edited by Paul Halmos, 93–100. Keele: University of Keele.

Banta, Trudy W. 2010. "Impact of Addressing Accountability Demands in the United States." *Quality in Higher Education* 16 (2): 181–183.

Barker, Katharine. 2007. "The UK Research Assessment Exercise: the Evolution of a National Research Evaluation System." *Research Evaluation* 16 (1): 3–12.

Beerkens, Maarja. 2015. "Quality Assurance in the Political Context: In the Midst of Different Expectations and Conflicting Goals." *Quality in Higher Education* 21 (3): 231–250.

Beerkens, Maarja. 2019. "Information Issues in Higher Education Markets." In *Encyclopaedia of Higher Education Systems and Institutions*, edited by Pedro Teixeira, Jung Cheol Shin, A. Amaral, A. Bernasconi, A. Magalhaes, B. M. Kehm, B. Stensaker, et al., in print. Dordrecht: Springer.

Beerkens, Maarja, and Maiki Udam. 2017. "Stakeholders in Higher Education Quality Assurance: Richness in Diversity?" *Higher Education Policy* 30 (3): 341–359.

Bowen, William G., Matthew M. Chingos, Kelly A. Lack, and Thomas I. Nygren. 2014. "Interactive Learning Online at Public Universities: Evidence from a Six-Campus Randomized Trial." *Journal of Policy Analysis and Management* 33 (1): 94–111.

Bramming, Pia. 2007. "An Argument for Strong Learning in Higher Education." *Quality in Higher Education* 13 (1): 45–56.

Brennan, John, and Tarla Shah. 2000. *Managing Quality in Higher Education: An International Perspective on Institutional Assessment and Change*. Maidenhead: Open University Press.

Brown, Roger. 2013. "Evidence-based Policy or Policy-Based Evidence? Higher Education Policies and Policymaking 1987–2012." *Perspectives: Policy and Practice in Higher Education* 17 (4): 118–123.

CACREP (Council for Accreditation of Counseling & Related Educational Programs). 2017. Accessed April 24, 2018. www.cacrep.org.

Carini, Robert M., George D. Kuh, and Stephen P. Klein. 2006. "Student Engagement and Student Learning: Testing the Linkages*." *Research in Higher Education* 47 (1): 1–32.

Coates, Hamish. 2005. "The Value of Student Engagement for Higher Education Quality Assurance." *Quality in Higher Education* 11 (1): 25–36.

Damian, Radu, Josep Grifoll, and Anke Rigbers. 2015. "On the Role of Impact Evaluation of Quality Assurance From the Strategic Perspective of Quality Assurance Agencies in the European Higher Education Area." *Quality in Higher Education* 21 (3): 251–269.

DETAG (Department of Education and Training, Australian Government). 2015. *Higher Education in Australia: A Review of Reviews from Dawkins to Today*. Canberra: DETAG.

Dill, David D. 1999. "Academic Accountability and University Adaptation: The Architecture of an Academic Learning Organization." *Higher Education* 38 (2): 127–154.

Dill, David D. 2010. "We Can't go Home Again: Insights from a Quarter Century of Experiments in External Academic Quality Assurance." *Quality in Higher Education* 16 (2): 159–161.

El-Khawas, Elaine. 2010. "The Teacher Education Accreditation Council (TEAC) in the USA." In *Public Policy for Academic Quality: Analyses of Innovative Policy Instruments*, edited by David D. Dill, and Maarja Beerkens, 37–54. Dordrecht: Springer.

ENQA (European Association for Quality Assurance in Higher Education). 2015. *Standards and Guidelines for Quality Assurance in the European Higher Education Area*. Brussels: ENQA.

ENQA (European Association for Quality Assurance in Higher Education). 2016. *Report of the ENQA Working Group on the Impact of Quality Assurance in Higher Education*. Brussels: ENQA. Accessed April 24, 2018. http://www.enqa.eu/wp-content/uploads/2016/05/Impact-WG-Final-Report.pdf.

Even, Trigg A., and Chester R. Robinson. 2013. "The Impact of CACREP Accreditation: A Multiway Frequency Analysis of Ethics Violations and Sanctions." *Journal of Counseling & Development* 91 (1): 26–34.

Ewell, Peter T. 2010. "The US National Survey of Student Engagement (NSSE)." In *Public Policy for Academic Quality: Analyses of Innovative Policy Instruments*, edited by David D. Dill, and Maarja Beerkens, 83–97. Dordrecht: Springer.

Faber, Marike, and Jeroen Huisman. 2003. "Same Voyage, Different Routes? The Course of the Netherlands and Denmark to a 'European Model' of Quality Assurance." *Quality in Higher Education* 9 (3): 231–242.

Gordon, Jonathan, Joe Ludlum, and J. Joseph Hoey. 2008. "Validating NSSE Against Student Outcomes: Are They Related?" *Research in Higher Education* 49 (1): 19–39.

Harvey, Lee. 2006. "Impact of Quality Assurance: Overview of a Discussion Between Representatives of External Quality Assurance Agencies." *Quality in Higher Education* 12 (3): 287–290.

Harvey, Lee. 2016. "Lessons Learned from Two Decades of *Quality in Higher Education*." Draft. Accessed April 24, 2018. https://www.qualityresearchinternational.com/Harvey2016Lessons.pdf.

Harvey, Lee, and James Williams. 2010. "Fifteen Years of *Quality in Higher Education*." *Quality in Higher Education* 16 (1): 3–36.

Head, Brian W. 2008. "Three Lenses of Evidence-Based Policy." *Australian Journal of Public Administration* 67 (1): 1–11.

Holcomb, Cheryl, Julia Bryan, and Stephanie Rahill. 2002. "Importance of the CACREP School Counseling Standards: School Counselors' Perceptions." *Professional School Counseling* 6 (2): 112–119.

Hopkin, Antony Gerald, and M. B. Lee. 2001. "Towards Improving Quality in 'Dependent' Institutions in a Developing Context." *Quality in Higher Education* 7 (3): 217–231.

Hopkins, David S. P. 1990. "The Higher Education Production Function: Theoretical Foundations and Empirical Findings." In *The Economics of American Universities: Management, Operations, and Fiscal Environment*, edited by Stephen A. Hoenack, and Eileen L. Collins, 11–32. Albany, NY: State University of New York Press.

Houston, Don, and Ahmed Ali Maniku. 2005. "Systems Perspectives on External Quality Assurance: Implications for Micro-States." *Quality in Higher Education* 11 (3): 213–226.

Klein, Stephen, Roger Benjamin, Richard Shavelson, and Roger Bolus. 2007. "The Collegiate Learning Assessment." *Evaluation Review* 31 (5): 415–439.

Leiber, Theodor, Bjørn Stensaker, and Lee Harvey. 2015. "Impact Evaluation of Quality Assurance in Higher Education: Methodology and Causal Designs." *Quality in Higher Education* 21 (3): 288–311.

Liu, Shui-Yun, Minda Tan, and Zhao-Rui Meng. 2015. "Impact of Quality Assurance on Higher Education Institutions: a Literature Review." *Higher Education Evaluation and Development* 9 (2): 17–34.

Marincovich, Michele. 2007. "Teaching and Learning in a Research-Intensive University." In *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*, edited by Raymond P Perry, and John C. Smart, 23–37. Dordrecht: Springer.

Massy, William F., and Nigel J. French. 2001. "Teaching and Learning Quality Process Review: What the Programme has achieved in Hong Kong." *Quality in Higher Education* 7 (1): 33–45.

Meyer, J. H. F. 1999. "Variation and Concepts of Quality in Student Learning." *Quality in Higher Education* 5 (2): 167–180.

Minelli, Eliana, Gianfranco Rebora, Matteo Turri, and Jeroen Huisman. 2006. "The Impact of Research and Teaching Evaluation in Universities: Comparing an Italian and a Dutch Case." *Quality in Higher Education* 12 (2): 109–124.

Neumann, Ruth T., and James Guthrie. 2009. "Performance Indicators in Australian Universities: Establishment, Development and Issues." Presented to the EIASM 2nd workshop on the process of reform of university systems, Venice, Italy, May 4–6, 2006. Accessed April 24, 2018 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1361320.

OECD. 2013. *Assessment of Higher Education Learning Outcomes: Feasibility Study Report. Volume 1: Design and Implementation*. Paris: OECD.

Pascarella, Ernest T. 2001. "Identifying Excellence in Undergraduate Education *Are We Even Close?*" *Change: The Magazine of Higher Learning* 33 (3): 18–23.

Rosa, Maria João, Diana Tavares, and Alberto Amaral. 2006. "Institutional Consequences of Quality Assessment." *Quality in Higher Education* 12 (2): 145–159.

Shavelson, Richard. 2009. *Measuring College Learning Responsibly Accountability in a New Era*. Palo Alto: Stanford University Press.

Skidelsky, Robert. 1992. *John Maynard Keynes: The Economist as Saviour, 1920–1937*. London: Allen Lane.

Solesbury, William. 2001. "Evidence-based policy: Whence it Came and Where it's Going." ESRC UK Centre for Evidence Based Policy and Practice: Working Paper 1. Queen Mary University of London. Accessed April 24, 2018. https://www.kcl.ac.uk/sspp/departments/politicaleconomy/research/cep/pubs/papers/assets/wp1.pdf.

Stensaker, Bjørn. 2003. "Trance, Transparency and Transformation: The Impact of External Quality Monitoring on Higher Education." *Quality in Higher Education* 9 (2): 151–159. doi:10.1080/13538320308158

Stensaker, Bjørn. 2007. "Impact of Quality Processes." In *Embedding Quality Culture in Higher Education. A Selection of Papers from the 1st European Quality Assurance Forum*, edited by Lucien Bollaert, Sanja Brus, Bruno Curvale, Lee Harvey, Emmi Helle, Henrik Toft Jensen, Janja Komljenovič, Andreas Orphanides, and Andrée Sursock, 59–62. Brussels: European University Association.

Stensaker, Bjørn. 2008. "Outcomes of Quality Assurance: A Discussion of Knowledge, Methodology and Validity." *Quality in Higher Education* 14 (1): 3–13. doi:10.1080/13538320802011532

Stensaker, Bjørn, Liv Langfeldt, Lee Harvey, Jeroen Huisman, and Don F. Westerheijden. 2011. "An in-Depth Study on the Impact of External Quality Assurance." *Assessment & Evaluation in Higher Education* 36 (4): 465–478. doi:10.1080/02602930903432074.

Suchanek, Justine, Manuel Pietzonka, Rainer H. F. Künzel, and Torsten Futterer. 2012. "The Impact of Accreditation on the Reform of Study Programmes in Germany." *Higher Education Management & Policy* 24 (1): 1–24.

Westerheijden, Don F. 2007. "States and Europe and Quality of Higher Education." In *Quality Assurance in Higher Education: Trends in Regulation, Translation and Transformation Higher Education Dynamics*, edited by Don F. Westerheijden, Bjørn Stensaker, and Maria João Rosa, 73–95. Dordrecht: Springer.

Westerheijden, Don F., Bjørn Stensaker, and Maria João Rosa, eds. 2007. *Quality Assurance in Higher Education: Trends in Regulation, Translation and Transformation Higher Education Dynamics*. Dordrecht: Springer.

White, Howard. 2009. "Theory-based Impact Evaluation: Principles and Practice." *Journal of Development Effectiveness* 1 (3): 271–284. doi:10.1080/19439340903114628.

Young, Ken, Deborah Ashby, Annette Boaz, and Lesley Grayson. 2002. "Social Science and the Evidence-Based Policy Movement." *Social Policy and Society* 1 (3): 215–224.