Knowing and Doing Vocational Education and Training Reform: Evidence, Learning and the Policy Process

Simon McGrath; Rosemary Lugg

Abstract: Much of VET policy internationally draws on a toolkit that has been seriously questioned for its logic, international relevance and effectiveness by considerable amounts of academic research. Reflecting primarily on our experiences of leading a complex, multi-country policy study, we develop an account that seeks to explore ways in which the apparent incommensurability between academic and policy knowledge can be addressed. This leads on to a broader discussion of key issues of contestation in the debates about knowledge for policy as they relate to international education and development more generally. We consider three key turns in the discourse of international education policy and research: to "governing by numbers", "what works" and policy learning, and ask what happens when these discursive trends travel to Southern and VET contexts. We suggest that this analysis implies that policymakers need both to be more modest and reflexive in their expectations of what knowledge can be mobilised for policy purposes and more serious in their commitment to supporting the generation of the types of knowledge that they claim to value. For international and comparative educators, we stress the importance of being clearer in seeking to shape research agendas; more rigorous in our approaches to research; and better in our external communication of our findings. Given the particular focus of this special issue on VET, we end by reiterating the particular challenge of reawakening research on VET-for-development from twenty years of slumbers.

INTRODUCTION

There is typically a large gap between policy and academic accounts of vocational education and training (VET). Several articles in this special issue refer to the VET policy toolkit, which is "known" by policymakers to be a sound basis for their work as such reforms have worked elsewhere and so will work in their countries too. However, these articles go on to present a very different set of accounts in which academics have a different knowledge, namely that these travelling policies exist more soundly at the level of discourse than that "of objective fact" and / or national contexts matter more than generalisable laws of VET reform. In this view, the toolkit does not work, or at least not fully and consistently. Indeed, one of the other papers presented in the UKFIET conference strand from which this special issue derives, provided an exemplary case from Mozambique. In it, Billetoft (2011) showed how, with donor support, Mozambique constructed a best practice policy but has not been able to make it work because it is not grounded in a realistic sense of how policy can be operationalised in this particular national context.[1]

Beyond the specifics of VET policymaking, there are wider processes at work in education policymaking globally and in development planning that are seeking not only to build policy on a stronger evidential base but also to shift academic research so that it is more relevant and responsive to the needs of policymakers. Although such approaches have been around in both education and in other fields of public policy in some developed countries for as much as a quarter of a century, there appears to be a rapidly rising interest amongst development cooperation agencies regarding such approaches as they apply to education-for-development. These trends are too new to have attracted much in the way of academic analysis and this is an important motivation for the current article.

---

[1] See also Norrag News 46 for a discussion of many of the issues of this volume.

For academics working in the field of international and comparative education (the core readership of IJED), notions that deconstruct international educational best practices are commonplace (cf. Chisholm and Leyendecker, 2008; Schweisfurth, 2011), even if the specificities of the evidence-based turn are not yet widely analysed. However, here we want to go beyond the current debate in two significant ways. First, we want to give more attention to addressing the apparent non-commensurability of academic and policy knowledges, using the specific area of VET as a way to stimulate a broader debate in the field of education-for-development. Second, we want to move past academic ways of knowing and doing in order to explore the possibilities that may lie in policy research for building bridges between policy and research as sites of knowledge production and practice. Of course, these bridges may be constructed differently to those built through the approaches to research deployed by policy analysts.

In this paper, we offer our reflections of collaborative work in a single complex multinational study, whilst drawing also on a combined nearly 40 years' worth of work in academic, consultancy and policy settings on three continents. We seek to engage in a debate between academic and policy approaches rather than simply showing why policy cannot work, as has been the tendency of much of the academic work on this topic. In doing so, we will reflect both on what we consider to have been an important attempt to produce better policy at a regional level and on some of the wider methodological and epistemological challenges that exist in international policymaking, using the VET sector as a particular example.

We will consider how our experiences inform three key debates about the knowledge for policy as it exists in the education sector. First, we consider the turn to "governing by numbers" (Rose, 1991; Ozga, 2009) and ask what happens when this discourse travels without the accompanying baggage of the technical and social capacities that are central to its operation in countries such as England. Second, we explore the way that a spreading to education-and-development settings of the discourse of "what works" research (Glass, 1987) and evidence-based policy (Thomas and Pring, 2004) is problematic and appears to contradict another element of the aid-development toolkit of the past decade – the move to working primarily at the policy level. Third, we interrogate the rise of policy learning as a potentially more emancipatory and egalitarian attempt to break out of policy borrowing, and as a practice through which to articulate policy research within policymaking, following and expanding upon Raffe's (2011) critique of the limits and possibilities of policy learning in a Scottish context. These three debates overlap considerably and this will lead us to come back to the same issues and examples across the three themes in a way that is iterative and respectful of the messiness of reality.

An exploration of these debates leads us to argue further that another major issue lies in the dominant tacit theory of policy both as being evidence-based and a driver of change. All of these issues are underpinned by two questions that prompted our writing of this article: why is there such a large gap between the policy stories about what works in VET internationally and the academic stories of what doesn't work? And what, if anything, can be done to bridge this gap?

We are very conscious and critical of the wider tendency to impose Northern theory on the South (cf. Connell, 2007; McGrath and Mason, 2011), so it is important to acknowledge the use of these Northern accounts, and broader Northern literatures on social theory and on policy that underpin such accounts. Our intention is not to see these theories as universal and then apply them to Southern contexts. Rather, it is to use them as a set of lenses through which we can set up a tension between global policy discourses and complex national and regional spaces in which practice, policy and research are enacted.

SOME METHODOLOGICAL CONSIDERATIONS

The writing of this paper has been challenging for two interlocking reasons that are important to make more transparent to readers. First, we are writing as two authors but the main project we are reflecting upon included 20 researchers, two funding agencies and several hundred stakeholders who participated in national and regional workshops. This raised important ethical questions regarding our right to write a single version of the multiple accounts that could be told of that project, particularly as European-based participants in a process that was largely African in its ownership and participation. We concluded that all we could do was to be respectful of the roles of others and mindful of the potential partiality and partialness of our account. Crucially, we are convinced of the importance of what we want to say and that it would remain unsaid if we do not take advantage of our capabilities to publish in an international journal.

Second, we are grappling with serious epistemological and methodological challenges. On deciding to write this paper, we knew that there were very major theoretical issues with which we wanted to engage based on a particular project in which we had collaborated but also arising out of our much larger combined experience of professional experience straddling academia, consultancy and policy work. However, we also knew that our knowing came from our professional engagement in the world and our critical reflection thereon rather than primarily from systematically collected data. Also, we were very aware that the impossibility of triangulating much of what we thought we knew against the perceptions of others could make our story open to contestation as codified knowledge. Moreover, some of what we knew was also potentially unsayable in other ways: because of confidentiality clauses signed in some of the many projects in which we have worked; or because of the challenges of maintaining access to potential clients and funders. Nonetheless, it is our contention that these matters needed to be inserted into the academic debate, particularly as this special issue has a key question running through it to which we could contribute. Nagging away at both of us, but also other authors in this special issue, is the question of why there is a VET orthodoxy that seems to be oblivious to much of the evidence about its limited effectiveness. If the challenges and dangers of working across the research and consultancy domains are contributing to this, then we wished to engage in this discussion. Moreover, we realised that we needed to do so in a way that didn't suggest that policymakers are naïve about research, but also attempted to provide at least some critical reflection on the weaknesses of academic approaches. We were also concerned to see if there was an insurmountable gap between different ways of knowing and doing VET policy and research or whether the potential existed for bridging this gap.

What follows has three parts that are intended to bring differing lenses to bear on a complex and multifaceted issue. First, we introduce and discuss the project that led to us writing this paper, exploring it as a case of trying to bridge the gap we identify above. Second, we seek to problematise this case (and some of our other experiences in policy and evaluation work) within a wider theoretical discussion of the challenges that emerge from thinking about processes of learning and evidence gathering for policymaking. Third, we conclude with a reflection on what the interplay between these two levels of discussion provides in the way of some possible conclusions about the limitations and possibilities of more effective dialogues between evidence and policy, and between academics and policymakers.

THE SADC/UNESCO PROJECT
Purposes of the SADC/UNESCO project

In 2009 SADC announced that it was going to embark on a major policy review of VET in the Southern African region, as the prelude to a new regional policy initiative in this area. This study was to be co-financed by UNESCO. The origins of this announcement

appear to come from three sources. SADC had produced a regional protocol on VET in 1997 (SADC, 1997) and had followed this up with work both on qualifications frameworks and on quality assurance systems (SADC, 2005 and 2007). However, there was a concern that a broader VET transformation agenda was important if the regional protocol was to translate into significant and widespread impact. As has been noted elsewhere in this special issue (McGrath, 2012), VET has become a more significant policy concern for UNESCO in recent years, and the proposed collaboration with SADC fitted very clearly to UNESCO.s preference for working collaboratively on policy work with national and regional structures. Moreover, for several years the CEOs of the region's VET agencies had been seeking to build a collaborative learning project that could go beyond earlier work on VET in some countries in the region through a British-South African collaboration (Akoojee, Gewer and McGrath, 2005). Thus, the project was built, at least in part, on interest amongst national, regional and international policy actors in cross-border policy learning and a desire to ground this learning in evidence about policy effectiveness and system development.

The eventual commissioning of this work also illustrated the concerns of these partners to work across the boundaries of academia, consultancy and policy. Although formally the heart of the project was a consultancy contract between SADC and a Netherlands-based consultancy firm, part of the attraction to the commissioning partners was that the team was led by a professor, who had also been the project manager of the earlier regional study and was widely published in this area, and included staff based at a regional university who had also previously worked for a national qualifications authority. The assessment and review of VET in the SADC region was implemented through a partnership between SADC, UNESCO and 13 of SADC.s member states and was intended to achieve several objectives at both national and regional levels. One of the key purposes was to build a regional strategic response for the revitalisation of VET in the region. This was to be founded on a baseline survey of the status of VET in the region, and the identification of intervention areas that would be "levers for VET in the region" (SADC and UNESCO, 2009). The logic of the intervention, therefore, was that comparative research on VET systems could identify the main drivers for reform, and that these could provide the basis of regional strategy to develop VET. This appears to have been grounded in an implicit view of reform that it is a technical-rational process that is driven from above, and one in which local subjectivities and contexts are irrelevant. We will come back to this issue later in the article.

The implementation of the baseline survey of VET in the region was associated with a second purpose for the project; the development of a monitoring tool that would allow SADC member states to track and report their advancement in VET and milestones achieved (SADC and UNESCO, 2009). The tool was intended to "allow countries to identify peers in similar stages of implementation of various aspects of their reform processes allowing for more coherent and coordinated learning opportunities and discussion of common challenges" (SADC and UNESCO, 2009:4). Therefore, it would serve the dual purposes of monitoring progress in national VET reforms, as well as supporting peer learning through regional exchange, particularly the discussion of best practices and constraints identified through the review. Although the terms of reference did include a concern with constraints, this appeared in practice to be less important than best practices. Thus, the drive was to help countries identify "what is working" and, potentially to consider "why", or "what makes it work"?

Therefore, the project brought together ideas of regional monitoring of national VET system development, identification of best practices and peer learning through regional exchange. These continue to be central themes within discussions about VET in the SADC region (UNESCO-UNEVOC, 2011).

Although the purpose of the project was to support policy learning within VET, the project was implemented through consultancy contracts with two teams of consultants:

a team of regional consultants, and 13 sets of national consultants. Although contracted separately, the regional and national teams were closely interdependent. The regional team comprised of five part-time team members, responsible for developing the research methodology and tools; providing research support to national consultants; analysing national findings and writing the regional report; and translating national experiences into a coherent regional strategic framework. National consultants were contracted by the UNESCO Cluster Offices in the region, and selected and appointed in each country by the relevant UNESCO National Commission. The UNESCO-appointed national consultants were all experts in VET, with particular expertise concerning the national VET context, and professional authority to lead a national, consultative study. Across the project, national consultants were drawn from a range of institutional locations, including freelance consultants, office bearers in national agencies, and university-based researchers. Many had prior experience as VET policymakers. Again, this reinforced the boundary crossing nature of the process, where researchers were selected because of their perceived ability to engage across academic, consultancy and policy domains. How these individuals knew VET in multiple ways, thus, was crucial and had major implications for how the project worked, as we shall consider later.
The national consultants had primary responsibility for implementing the study within a specific national context, for data collection and reporting, and for facilitating a national stakeholder workshop to validate the national VET monitoring report and to identify national strategic priorities for regional support.

Developing the tool: indicators

Drafting the monitoring tool was the responsibility of the regional consultants. Although neither SADC nor UNESCO sought to specify the form that the monitoring tool should take, key areas to be addressed by the tool were identified. These included:

. VET policy frameworks;

. size, shape and structure of the system;

. governance;
. financing;
. management and information systems; and
. teacher training.

SADC was also concerned to determine the extent to which there was a tendency towards harmonisation of VET policies, qualifications and quality assurance systems in the region. These are familiar features that are also considered through established monitoring processes in other regions including ReferNet in Europe. Indeed, the rise of approaches to cross-national comparison can be considered to be an important element of the international VET toolkit and, indeed, of wider policy tool kits linked to the rise of governance and performativity.

The range of areas of interest suggested monitoring through different kinds of indicators and different forms of reporting data. For this, both qualitative and quantitative data were necessary, and therefore a mixed methods approach to the tool was proposed by the regional consultancy team, and agreed by the funders. Again, this is a decision that was of great significance in shaping how the project came to know VET in the region and provides a series of insights that we will bring to the wider discussion of VET knowledge for policy below.

Two considerations were critical in the development of indicators for the monitoring tool. First, as dutiful consultants, we were strongly influenced by the priority areas set out in the terms of reference (see above). Second, we shared with SADC and UNESCO an

awareness that similar processes were ongoing elsewhere internationally. Indeed, the development of regional indicator sets for VET is most developed in the European Union. Our review of examples of indicator / cross-national reporting development in the EU was practically mindful of the very different contextual situation in Europe, including matters of capacity and resourcing, which meant that European models could not be borrowed for the SADC context. This reading was also theoretically informed by our grounding in the Sadlerian comparative education tradition, with its very strong concerns regarding matters of context (Crossley and Jarvis, 2001).

In addition to reviewing monitoring tools used to evaluate European systems, work within the SADC region was also considered. This provided a sense of what data might be available at the national level in member states. Inevitably, this showed that there was a massive mismatch between research, monitoring and evaluation capacity in South Africa and in the rest of the region.

A very provisional series of indicators and/or descriptors that the monitoring tool could include was identified. In considering indicators, there was a clear tension between a necessary ambition to cover as wide a range of data as might be seen to be required to give a full picture of a VET system, and an equally necessary modesty regarding what was feasible in terms of existing data sources and the available capacity for national officials to gather the requested data in the future. It was realised that some indicators may have been considered so important that they should be included even though data collection currently would be difficult. Issues of what data can be found readily and the dangers of being driven by data availability rather than fundamental policy goals will be returned to in the next section of this article.

More than 100 possible indicators were identified. These were reduced to a list of 36 based on an analytical reading of their importance for the process of VET reform in Southern Africa. This list was then ranked according to a combination of the priority of the indicator and the likely availability of data. Through dialogue with SADC and UNESCO, the long list of 36 was reduced to a second list of 18 indicators. Although this list was considered to be rather long, and we offered a further ranking of them against the above criteria of priority and availability, SADC and UNESCO officials felt strongly that all 18 indicators would be piloted as a strategy to test the limits of the tool. These indicators were then refined through a workshop with the national and regional consultants and both commissioning agencies, and the exact nature of the data to be collected against each agreed (the indicators are listed in Appendix 1). The tool included indicators of context, input, process, output and outcomes. Three broad kinds of indicator and report were included:

• narrative reports,
• ordinal reports (based on criterion-referenced ordinals) and
• statistical reports.

Whilst all indicators in the SADC monitoring tool are designed to assess the status of VET system development, ordinal reports provided a particularly immediate form of assessment. For these indicator reports, a five point scale was constructed with a score of 1 on the scale being associated with challenges faced without systemic interventions, through three broad phases of systemic development (scores 2, 3, 4), with a score of 5 reflecting a mature or strong performance in terms of the indicator. A concern in developing the ordinal indicators was managing assumptions implicit within such scales of a pre-determined, ideal form of system development. We were also concerned about how ratings awarded within national relative frames of reference might become translated, inappropriately, into regional rankings. Again, these are issues to which we will return in the next part of this article.

Piloting the monitoring tool

National consultants carried out the data collection required to complete the draft national VET monitoring report. This focused on accessing data that already existed at the national level, and three main sources of data were mined. To start with, the national consultants carried out an extensive review of the literature on VET in their own country. This included research literature, national policy documents and published reports on VET as well as any available monitoring or evaluation reports. National VET management and information systems were consulted in order to access data required to complete the reports on statistical indicators. Finally, a small number of interviews were carried out with key stakeholders in order to gain additional information required. The draft reports were then validated by national stakeholders through national workshops. These largely confirmed the draft reports, and in some cases were able to provide access to additional data. Significantly for the subject of this article, the workshops also revealed the value of the monitoring reports as a tool for dialogue within national VET systems. Stakeholders commented that they gained information on different parts of the sector and that debate around different perceptions on the system – perhaps particularly around different assessments and justifications for ordinal ratings – were extremely valuable.

This process also illustrated issues regarding objectivity and subjectivity of rankings; serious limitations in existing data and availability of key documents; and the weak state of national VET capacity in some countries: issues that raise significant questions for our subsequent discussions.

Outcomes of the project

Within a remarkably short period of five months, 13 national monitoring reports were drafted, validated within national workshops and finalised. A regional report was produced and agreed at a further regional workshop by the commissioning agencies and national stakeholders, as was a draft strategy for responding to its key lessons. This was then presented to the regional Ministers of Education and approved in September 2011. Whilst national stakeholders and national consultants broadly affirmed the value of the monitoring tool as a strategy to enable national stakeholders to reflect on system development, a number of concerns were raised, particularly around the feasibility of a single blueprint meeting the needs of all contexts, and regarding how key national priorities should be included. As the lead consultants on the project, we also believe that the tool and the project as a whole largely worked in spite of its complexity and short timescale. We are confident that the report did provide a stronger evidential base than was previously available from which to inform policy. Indeed, the SADC Ministers of Education accepted both the report and a set of policy recommendations from it. Thus, it would have been possible to present this story largely as a successful case of evidence-based policy and / or policy learning and to finish our account there. However, we want to go beyond this version of the story, grounded as it is in policy and consultancy ways of knowing. Our intention in what follows is not to say that this version of the story is not a reasonable one. Rather, we want to offer another way of knowing the same issues from an academic perspective that is more concerned with epistemological and methodological debates regarding the complexities of knowing and doing vocational education and training reform. In so doing, we will bring together our reflections on this specific project with our wider knowledge and experience gathered as academics, consultants and policy officials / advisors.

KNOWING AND DOING VOCATIONAL EDUCATION AND TRAINING REFORM: SOME KEY DEBATES
Governing by numbers

For the purposes of this paper, we want to look at three elements of this process of governing by numbers. First, there is a general point about the rise of statistics to govern the world (cf. Goldstein, 2008; Gorard, 2008; Ozga, 2009). Second, there is a more specific issue regarding the weakness of VET data, including problems of definition and multiple forms of delivery, which make any attempt at data generation even more fraught than developing indicators on schools (cf. Jansen, 2005; Gorard, 2008; Lewin, 2011). Third, there is a particular challenge regarding the objectification and quantification of qualitative and subjective judgments. Here, we will reflect on the specific case of our use of ordinal data in the SADC/UNESCO project but also some other experiences regarding how such ratings may change over time, a classic question of reliability.

From a social constructivist perspective, statistics cannot be understood apart from the professionals who develop and analyse them and the environments in which these actions take place.

Ozga (2009) critiques the ideological purpose of evaluation and statistics as arising from a shift from government to governance.

> The shift to governance is, in fact, heavily dependent on knowledge and information, which play a pivotal role both in the pervasiveness of governance and in allowing the development of its dispersed, distributed and disaggregated form. Data support and create new kinds of policy instrument that organise political relations through communication/information and hence legitimise that organisation (Ozga, 2009: 150)

Thus the rise of statistics can be linked to the metaphor of toolkits used in McGrath's and Lim's papers in this issue. In this case, there is a governance toolkit, consisting of notions of governance, performativity, decentralisation, marketisation, etc., that both sits behind the VET toolkit unseen and which erupts into the VET toolkit around issues such as decentralisation and governance reforms. Indeed, one of the largest and most interesting discussions within the SADC/UNESCO project was around the issue of decentralisation. Whilst clearly part of this toolkit, or set of travelling policies, linked to the new public management (cf. McGrath, 2010a), decentralisation of VET was not seen as an unquestionable good to which numerical value could be ascribed. Rather, for South Africa in particular, a national belief in the need for the state itself to be performative and developmental led to a strong view that decentralisation led to poor policy implementation and, therefore, undermined achievement of developmental goals.
At a methodological level, Gorard (2008) emphasises the professional judgments that typically lurk unseen behind statistics as publically presented. He reminds us that all statistics are subject to professional decisions regarding what to collect and how; how to manage data problems; and how to analyse and present data. This makes it impossible in strict terms to make confident comparisons of statistics over time and or space and undermines the processes of doing secondary analysis. Indeed, we would go further and argue that, epistemologically, those professional decisions are fraught with ideology and power.

In the context of EFA, Lewin (2011: 576) argues that the understandable drive towards indicators brings potential problems:

> the discourse around EFA uses a range of key indicators to define and monitor progress. The problem is that each of these indicators comes with characteristics that may not be fully understood by those that use them. The

line between simple indicators that have traction in the politics of aid and the pressure on development partners and governments to be seen to "do something that makes a difference" sits uneasily with targets and indicators that are meaningful to both target setters and target getters and which can be used flexibly across a range of contexts. There is also a tension between indicators which satisfy ambitions for precision and comprehensiveness, and the realistic prospect of having reliable data on which to base their realisation and use.

It was evident in the SADC/UNESCO process that there was a profound gap between our academic caution regarding the meaning of the statistics with which we were working and the concern of officials to have and use "hard" statistical data. The explanation for this may partly lie in different sources of risk. For the academic, a major source of risk comes from overclaiming and being too certain, given Popperian concerns about falsification and provisionality. However, for policymakers, and thus for consultants and advisors, risk lies far more in uncertainty and in the lack of clear and decisions and decisive actions.

If Gorard can raise the above concerns about interpretation of quantitative data in the statistically sophisticated setting of England, then it is easy to imagine how more problematic the challenge is likely to be in the far more resource-poor settings of Southern Africa. Thus, in the SADC/UNESCO study we were faced by very serious questions regarding the nature of statistical data. Moreover, we inevitably faced a delicate and political tension regarding decisions of whether ultimately to exclude data which our professional judgment told us were below our threshold of acceptability even where these were official data that national stakeholders "knew to be true".

Ozga raises a wider issue about the relationship between data, governance and the rise of networks. She argues that England has experienced a shift from a state that collects data to a web of educational institutions that both pass data upwards but also internally judge themselves against their own and others. data through a variety of proprietary software tools– holding heads, teachers and subject teams to account for below average student performance against objective benchmarks. Thus, in England, the profession becomes owned by the performativity agenda in its owning of and acting in response to this data.

However, the broader cultures and technologies of this approach are simply not available to SADC or African governments. Thus, weak statistics combine with underdeveloped state capacities and restricted institutions to undermine much of the key work that such statistics are supposed to do. Moreover, they exist in a particular socio-cultural milieu that also limits the desirability and effectiveness of such approaches. Governing by numbers in poorer countries thus seems doomed to failure at the present time.

. The unbearable lightness of VET data

It is widely accepted that VET data are weaker and more problematic than education data, both in the SADC region and globally. This is acknowledged in the UNESCO World Report on TVET (UNESCO, 2012), whilst a UNESCO regional report on Africa (UNESCOBREDA, 2010) argues that such data have got worse over recent years. The inadequacy of such data was also a key finding of our report and led to the regional Ministers agreeing to prioritise strengthening VET management and information systems, as part of a wider commitment to improved research, monitoring and evaluation.

To some extent, this reflects the low status of VET in recent years (cf. McGrath, 2011 and 2012). As Goldstein (2008) notes, decisions on what statistics to collect and with what vigour and rigour, is indicative of policy priorities, reflecting in turn wider societal debates and contestations in a process that is profoundly mediatised. In this, VET is not

a particular priority in most countries.

However, there are also particular problems in how to define VET (McGrath, 2012; UNESCO, 2012) that have profound implications for what it means to collect and analyse VET statistics. What counts as VET enrolments in official education statistics is based on counting those in vocational streams in upper secondary schools or in public vocational institutions under the control of the Ministry of Education. However, the increasing hybridisation between academic and vocational education (UNESCO, 2012) makes it difficult to know what should be included here and what has been included by Ministry officials across national settings. Excluded from such enrolment figures, however, are the far larger numbers of learners who are engaged in vocational learning in settings such as public vocational institutions under other Ministries such as Health and Agriculture; private vocational colleges; and formal and informal sector workplaces. Neither in the 13 countries studied as part of the SADC/UNESCO project or elsewhere is it possible to be confident about the numbers engaged in VET in a particular country, making international comparisons fundamentally impossible.

If enrolment data is fraught with serious problems, it can be imagined that there are even greater problems with notions of retention, success and destinations. Indeed, one of the most prolonged discussions in the SADC/UNESCO project design workshop was over the notion and measurement of success given that even public vocational systems were divided between examinations-and competency-based assessment; whilst some collected data by subject or module and others by qualification or candidate. Moreover, as with schooling, we were not confident that higher pass rates in VET, however defined, could be taken as an indicator of higher quality, being very mindful of the ways in which examinations results in particular are open to contestation and manipulation at various levels.

Understandably, however, the SADC/UNESCO report needed to provide some such statistics and so the focus became one of trying to ensure as much agreement as possible across countries as to what could be counted that might be comparable. Alongside this, we insisted that each country monitoring report should contain meta-data regarding what definitions of particular indicators were used; what specific problems were encountered with the data; and what possible absences and irregularities might exist in the numbers presented. These are all sensible attempts to respond to some of Gorard.s concerns but it is clear that there is a strong risk that such meta-data gets seen as interference, distorting the clarity of the policy signal that was the very intention of gathering the data. In response, it is likely that comparative reporting of data will tend to reify the figures and omit the caveats. Indeed, one of the tensions that ran through the project was between our concern to avoid the easy construction of league tables and a regional concern to have clear evidence for benchmarking country performance.

. The seductiveness of quasi-numbers

All participants in the SADC/UNESCO process were sure that the sum of VET reform could not be captured by statistics alone. However, there was some interest in being able to look at and compare simple scores for as much of what was being examined as possible. This led us to develop a set of ordinal ratings, as noted above, for themes such as quality assurance, management information systems and industry involvement in VET. This followed the approach of previous single issue SADC reports (SADC, 2005 and 2007).

In every case, we strove to ensure that these numbers were the starting point of a discussion of a topic in the national reports. Each ordinal rating was followed by a detailed narrative justification of the grade given. However, it is clear that there is a real danger here too that the complexity and nuancing of the message is likely to get lost in the face of a simple number.

Such quasi-numbers also bring particular issues regarding how they are constructed and how they are then treated as real numbers subject to the usual tools of mathematical manipulation, such as the creation of a national mean score across indicators or comparisons of mean ratings across indicators.

The creation of the numbers was a particularly complex process. We developed level descriptors for each ranking for each indicator and discussed these with national colleagues at the design workshop. However, we cannot be confident that we all understood these descriptors and the evidential requirements to justify particular ratings in the same way. Moreover, each rating was then subject to validation at national stakeholder workshops. There, a few ratings were adjusted and we have only limited access to the reasoning on which these adjustments were made (cf. Gorard's concerns about the limited transparency of professional decisions regarding statistics above).

By being constructed as numbers, these ratings took on an air of mathematical precision and scientific detachedness. However, it became apparent in our process of constructing a regional report that national reports reflected complex sociological and psychological processes. In a number of cases, our supposedly objective reading of the evidence as outsiders suggested a higher or lower ranking than that provided. In a number of cases, there appeared a tendency to consider that a policy being discussed was as significant as having a policy promulgated; and that a policy document was equivalent to a policy implemented with demonstrable evidence of outcomes and impacts. At least three things appear to be going on here. First, there was an optimism that the change process was simple and rapid. Second, there may also have been cases where political spin became a factor. Third, this relates to a policy turn in which international development cooperation has shifted from projects to policies. Whilst policies are supposed to be attractive because they ensure a greater scale of impact, this has also led to a new discourse in which the policy becomes everything. We will return to this in our discussion of "what works" below.

In the particular case of South Africa, a rather different dynamic was at play. South African self-ratings were often in the middle range of national ratings on the ordinal indicators. These seemed to be more realistic, perhaps even pessimistic, than was the case in other countries, with there being a number of areas where we would have rated South Africa higher if adjusting the ratings across countries. This could be reflective of the far higher national capacity in South Africa to analyse VET performance based on already-available research. However, there also seemed to be at least two important temporal effects at play. First, having had a national qualifications framework, for instance, for 15 years, it was unlikely that South African evaluators and stakeholders could be seduced into thinking that there was anything but a long road to impact. Second, the project was investigating South African VET at a moment when there was a major national process of policy review linked to a new government (see Akoojee, 2012 for more details). From Ministerial level, there had been a very clear message that key elements of the system were not working and this can be expected to have given informants permission to be critical.

Temporal effects are likely to become even more significant in this process if, as is planned, the monitoring tool is readministered on a periodic basis. In an earlier programme of evaluation of a large donor intervention into a national VET system, one of us had seen fascinating trends at work as evaluators revisited institutions every six months over a three year period. As part of this process, members of governing bodies, leaders and teaching staff were asked to rate a series of areas of institutional performance on a scale of 1-10. In several institutions these ratings started high, with respondents arguing that new policies and practices were rapidly being put in place. They tended to keep the rating high on the next visit, noting that these reforms were taking some time, but were nearly there. However, there were a number of cases in year two when ratings plummeted as staff became despondent about the pace of change and

the obstacles faced in realising institutional change.

"What works" and the search for evidence-based policy

For SADC and key regional stakeholders there was a strong sense that the importance of the exercise lay in finding out more about what worked in VET nationally, regionally and internationally. This knowledge could then drive improvement. Throughout the process, we were challenged to provide more evidence of what worked and of best practices. However, we constantly found ourselves making two responses that pointed to the impossibility of doing this. First, we were clear that the process could not generate robust conclusions about what worked as the evidential basis was too weak. There was a dearth of available evaluative evidence on interventions, and what evidence there was would not satisfy the usual requirements of the "what works" tradition, being poorly designed or very limited in scale and scope. Several of the reports were also reluctant to make such claims. Essentially, most of what was presented as best practice was assertions by officials and project leaders. Second, however, there was even very little of this "weak" evidence. As we noted above when discussing the ordinal ratings, it was quite common for policy documents to be seen as evidence of impact. Countries differed in the extent to which national consultants were able to communicate successfully that having a policy does not constitute evidence of having had a developmental impact. We shall turn to this second set of issues before returning to the first.

. How policy trumps impact

We can speculate that part of the problem here may have to do with a key tension in development discourse and practice. As we have noted, there was a policy turn in development, which took place from the mid-1990s. Yet, at the same time, there was a rise of meta-level developmental goals, enshrined most powerfully in the MDGs. These place the achievement of a small number of measurable goals at the heart of the development enterprise. In our experience, these two contradictory tendencies get connected together through a key tool of development practice: the logframe (and similar tools). Thus, in the large, longitudinal evaluation mentioned above, the principal focus of the programme was on transforming public vocational providers into versions of the English further education college, with a strong emphasis on governance, leadership and industry engagement. This is what the evaluation team were exploring in their six monthly visits and what the national government and the Chief Technical Advisor understood as the heart of the programme. However, when at the end of the programme, a group of external evaluators were sent from the donor country to examine the programme, their focus was back on the logframe in the initial project documentation and the stated overall aim of the project: to reduce poverty. How the programme was supposed to have moved from institutional governance reforms to poverty reduction was never particularly clear, and certainly not convincing, in the initial documentation but this had been elided by those involved in it who had seen it as engaged in the important business of VET reform. The national evaluation team had strong evidence of what had worked and not worked in the programme in terms of its core business; but the donor team struggled to find evidence that the programme had really worked in terms of its overall aim.

. Why "what works" won't work

However, beyond the rather soft concerns with what works and best practice in the SADC/UNESCO process there is a far harder edged drive emerging from a number of donors and the International Initiative for Impact Evaluation that seeks to export "what

works" approaches to social and economic development in the South.[2]

As it emerged in the North, particularly initially in the USA, this approach was subjected to a range of political, epistemological and methodological critiques. Glass (1987), for instance, argues that the original "what works" study in American education was part of a Reaganite project to wrest control of knowledge away from "dangerous" academics and to restore power to "right-thinking" people. He shows how the process was intensely political from the start without a rigorous methodology of proving what works. Areas where there was a lack of quantitative data couldn't be known about and some areas of heated contestation were similarly unknowable as there was no "what works" to be distilled. However, there was also a lack of an attempt at rigorous meta-analysis where this could have been possible.

Schoenfeld (2006) notes that the What Works Clearinghouse that built on this initial initiative had clear criteria for what research could be considered:

• Randomised experiments
• Quasi-experiments that use equating procedures
• Studies that use the regression discontinuity design

In a similar vein to Gorard's critique of statistics above, however, he argues that there is too much minute difference in studies to be confident in doing meta-analyses.

Schoenfeld's concern is that it is very difficult to do such work rigorously enough, but for many in the education-for-development community the concern is more an epistemological one. There are strong epistemological traditions in comparative education (Crossley, Broadfoot and Schweisfurth 2007) that question whether education can be known in the ways assumed by the "what works" approach, stressing instead the constructed, contingent and contextual nature of the social world.

However, it appears that this might be one of the most important challenges for international and comparative educators in the next few years. In a recent paper (McGrath, 2010b), one of us wrote of how we as a community needed to think carefully of why we were so marginal to debates about development when compared with economists. Our sense is that the rise of "what works" exacerbates this challenge and reinforces the need for critical reflection on whether we do want to try to influence policy and, if so, how we can best do this in the light of the power of other approaches and disciplines.

Whatever the merits of these critiques of the "what works" approach, at the present time it is simply impossible to apply this to the field of VET in Africa. Even in South Africa, it would be impossible to find a single study of VET that met the criteria Schoenfeld lists above. However, this contrasts with the rise in VET's importance in development thinking, especially on and in Africa, in the very recent past (cf. McGrath, 2011 and 2012). Thus, we are faced with a fundamental problem for evidence-based policy: VET interventions will be funded because there is political will but there is no evidence base on which to do this. This was the challenge that SADC and UNESCO faced in the project under consideration here but this led them quickly to a realisation that the way forward lay in broader and more flexible understandings of what counts as evidence.

As there is a turn back to VET in Africa, it seems vital that the question of evidence be taken seriously. As was noted above, the SADC Ministers of Education have agreed that better research, evaluation and management information systems are of strategic importance, and a similar argument is advanced by UNESCO in the World Report. There

---

[2] We are indebted to Kenneth King for reminding us that the World Bank showed an early interest in "what works" approaches to international education.

are good reasons why VET research in Africa is so weak at present (Carton, 2011). These would include familiar critiques of structurally-adjusted educational research; the low status of VET and hence its research; the negative effects of consultancy on research output; and the unwillingness of many implementers to generate rigorous and published evaluation. However, it is evident that the research community has a responsibility to show agency in responding to a new opportunity for rigorous VET research. Our epistemological position suggests that such research must use a variety of methods, and should not forget the importance of listening to the hitherto silenced voices of learners (Powell, 2012) alongside any rise of experimental methods.

Policy science for policy learning

Policy learning has become an important strand of thinking about the evidence-policy link in VET in recent years due largely to the efforts of the European Training Foundation (Grootings, 2004; Chakroun, 2008 and 2010).

As we have discussed, cross-border peer learning was one of the key purposes of the SADC/UNESCO project, and of the development of a monitoring tool in particular. As we noted much earlier in this article, monitoring tools have been used for policy learning purposes in Europe. ReferNet was introduced in 2002 to raise transparency in European VET and to improve knowledge about systems, to facilitate cross-border knowledge sharing amongst experts. Given the major differences between VET systems in Europe,

ReferNet was intended to "provide information to facilitate cooperation within a spirit of trust and to enable us to learn from one another" (Hippach-Schnieder, 2009: 28). The reports are intended to enable countries to monitor progress towards European VET policy priorities. According to the European Commission, the indicators are also to be used as a tool for understanding the reasons for differences in performance, and to support learning between countries on best practices (Dunkel, 2009).

Within the research literature, policy learning is distinguished from policy borrowing. Whilst both are voluntary and initiated by the recipient country, policy borrowing involves particular policies that a country seeks to imitate or copy from another (Dale, 1999; McGrath, 2010a). Policy learning, in contrast, is intended to be "a way for governments and systems of governance to inform policy development by drawing lessons from available evidence and experience" (Chakroun, 2008: 12). As well as drawing on the experiences of other countries, the evidence from which lessons are learned may be based on a country's own experience. Central to the idea of policy learning is that as a process it leads to policy change and increasing policy effectiveness (Grootings, 2004). This points to the need for evidence about policy effectiveness.

As McGrath (2010b) notes, the nature of the distinction between policy borrowing and policy learning can be rather subtle. He suggests that it is often "in the eye of the beholder" and may in the end be analogous to the difference between deep and surface learning (McGrath, 2010b). However, advocates of policy learning processes argue that differences are evident in the way that policy development processes are supported, facilitating inclusive learning processes, and evident in reformed policies that derive from, respond to and are driven by national contexts (Chakroun, 2010).

In a recent paper, Raffe distinguished between policy learning as "science" which "aims to identify valid policy lessons from other countries experiences" and policy learning as "sociology": "the learning that occurs, and the cross national influences on behaviour" (Raffe, 2011: 284). Raffe's distinction is valuable as it draws attention to the nature of the process of learning from experience. We understand his use of science to mean robust research, and as supporting research from a range of methodological positions, rather than a prioritisation of a particular form of scientific methodology, unlike the

"what works" approach.

In his paper introducing this distinction, Raffe demonstrates policy learning as science by carrying out a review of a particular policy (the NQF in Scotland) and considering to what extent, and why, the policy may be considered a success. In considering evidence to support both a "celebratory account" and a "skeptical account", and then comparing them, Raffe draws additional valuable lessons for policy development and offers us a deeper understanding of policy, its context and the nature and range of lessons that may be drawn therefrom. For Raffe, policy learning starts with researching experience, and reflecting on this within the practice of research. The lessons are drawn within the research domain. This corresponds to the deeper learning to which McGrath referred.

Possibilities for evidence-based policy learning about VET system development were limited within the SADC region by weaknesses in the VET knowledge base, as we have already noted. The lack of research is compounded by significant weaknesses in research dissemination, so that even research that has been carried out may not be widely known, and can perhaps be easily excluded from policy learning processes. This has already been recognised as an important contributing factor to a lack in policy learning in other regions (Chakroun, 2008).

As we noted earlier, VET statistics are made more problematic by the range and complexity of meanings and settings for VET. This also undermines policy learning as there is a weak co-construction of technical language available for cross-border dialogue. The construction of scientific accounts about VET, and in particular, comparative analysis of these accounts, is currently constrained by a lack of clarity about what VET is, and what the sector includes (and why). The development of a language through which dialogue about VET may be constructed, enacted, and communicated may imply the need for a taxonomy for VET (Shoesmith, 2011), although such a task is challenging.

The SADC monitoring tool provided a standardised set of indicators, sub-indicators and reporting formats through which different VET systems could be consistently described and assessed. Even so, the reports themselves reveal that these could not be standardised but were constructed within a series of dimensions, or tensions. These are not discrete categories, but the polarities represent a heuristic device for exploring how evidence-based accounts are also constructed in relative terms. The differential constructions of reports make it impossible to learn from cross-national experiences by simply reading across reports or making superficial comparisons between accounts in different reports.

As accounts of different systems are constructed differently in relation to each dimension, part of the complexity in cross-country policy learning derives from different constructions of research, of evaluation or of monitoring. However, exploring the tendencies within which reports are developed may open up methodological discussions that will strengthen both the science of policy learning and shared understandings of what counts as "evidence".

• research knowledge and political knowledge

Using Raffe's lenses, developing an account of a VET system is both a scientific process of collecting data through robust research methods, analysis and reporting, and a political process. Scientific validity of the account rests on the quality of the research process used and the quality of the data gleaned. Political validity is also built through particular methods, which may be contextually dependent. This is evident, for example, in practices such as the requirement to sign off reports at political levels, through a particular political process to validate data, in the institutional home from which the

report is written, and to which it is tabled. It has implications for what may be reported, by whom, and when.

Both forms of knowledge were essential within the SADC/UNESCO project, were deployed in constructing the validity of the research, and needed to be managed during the process. This is not unique to this project. For instance, difficulties in reconciling political and research processes have been managed in the German ReferNet reports through the use of two distinct sections: one completed by the Federal Ministry of Education and Research, and the other summarising academic research and other data (Hippach-Schneider, 2009).

• policy and performance

This tension refers to a tension between policy claims and system performance. As we have noted, the dearth of data, combined with the discursive turn to policy, leads to cases where policy eclipses performance to such an extent that having a policy vision becomes conflated with performance. In the absence of data on performance, policy learning is replaced by the exchange of policy claims. Allais points to problems generated by a lack of evidence of policy performance in the case studies of several NQFs recently carried out by the ILO (Allais, 2010).

• "outsider" and "insider" perspectives

Reports of VET systems may be constructed as written from outside the system, as objective observations; or they may be presented as self-assessments from inside the system. In the SADC/UNESCO project, possibilities for being outside a national system were limited, particularly in the small states that amount to half of the member states. National researchers may have been engaged from a range of positions, for example within a national think-tank, or as consultants, but these are not necessarily associated with objectivity. In constructing accounts, tensions exist between the status of objective outsider and knowing insider. At the regional stakeholder workshop towards the close of the project, the delegates proposed that future iterations of the monitoring process should be housed within VET systems so that reporting was part of a national self-assessment process. This reflects a significant move away from the pilot where national reports were filtered through the editorial lens of external consultants and, therefore, potentially capable of greater distancing from issues of local politics. If desired, this would need to be mediated within regional structures or processes.

• "positive" and "critical" accounts

We have already referred to our observation that some ordinal ratings and narrative accounts appeared to be optimistic or pessimistic, whilst Raffe used the terms celebratory and skeptical. What is driving both distinctions is a concern that there are different presentation styles through which data can be refracted, giving potentially very different interpretations. The construction of a report as positive or critical can be a matter of tone, but is also reflected in the selections – and omissions -of data, or particular research studies, and in the construction of the argument made based on the data reviewed. Indeed, it is this dichotomy to which Raffe points us, and uses productively in his demonstration of the science of policy learning.

Certain methods, such as stakeholder interviews, may also shape the overall shape of the data and therefore whether optimistic or pessimistic readings are available. As we noted above, our experience across several projects suggests that increasingly pessimistic accounts may emerge over time. This makes longitudinal policy learning

particularly challenging. We have also noted that the political context may shape the possibility for critical accounts. In some contexts it is not politically acceptable to be critical or optimistic. Thus, further temporal, but also spatial, effects enter into the analysis of what can and should be learned.

Finally, as is often the case, perceptions about the report's audience may shape the nature of the report. Concerns that reports may be used for purposes other than those intended by the respondents can be influential. For example concerns about a critical external gaze may reinforce a tendency towards optimism. Indeed, it was interesting to note in the SADC/UNESCO project that national stakeholders were often caught between a desire to be honest about their system's failings in their own interventions and a tendency to be defensive if others seemed to be drawing negative lessons from their systems. Learning is always fraught with dangers and policy learning seems to lead to particular risks that need to be managed by participants. However, their management strategies may also limit or distort peer learning.

• purposes for cross-national peer learning and for accountability

The language of peer learning is very different from the critique of governing by numbers above. However, in practice the "terrors of performativity" (Ball, 2003) are never far from the surface in policy learning. Thus, although the monitoring report was expressly couched in terms of peer learning, it was evident that there were concerns that this could also be a tool for monitoring performance. In the OECD and EU, peer reviews exert influence through peer pressure although this is still often weak when compared to the pressure exerted through international rankings, such as PISA (Hippach-Schneider; Dunkel; Speer, all 2009). In the SADC/UNESCO project, we were very resistant to what we perceived as a pressure from some participants to produce a series of what could be read as league tables, but the data and analysis clearly has a life beyond our control.

More positively, we see merits in stressing that policy learning is both science and sociology. It does have the potential to build a more collaborative model of policy making, including a wider range of stakeholders and through more democratic processes, thus changing the way policies are made (Raffe and Spours, 2007; Chakroun, 2008). It can also lead to the generation of more robust analyses of multiple forms of data.

However, this needs to be balanced against the way that travelling policies such as the VET and governance toolkits shape discursive possibilities in countries and regions. It must also be remembered that national policy learning is undermined internally in many fragile states or in cases of acute aid dependency. To the latter, we can add in practice the context of EU accession states. For these countries (at least before the crisis afflicting the EU as we write) becoming part of the EU is all that matters and there is little point in wasting time in policy learning about aspects of VET when the policy imperative is to sign up to everything that the EU considers best practice in this sector and others.

An implicit theory of policy

What appears as a key tension in these accounts, but also in the SADC/UNESCO project, is a theory of how policy works. Whereas the policy learning account sees policymaking as arising from a public dialogue about what works; the other strands discussed above are grounded in a far narrower technical understanding of both evidence and debate. However, in all these approaches there appears to be a view that good policy decisions can and will be made through technical or dialogic processes. Yet, of course, this stands

in stark contrast to a lengthy tradition in thinking about educational policymaking that is informed by sociological and political economy traditions. Here, policy is often understood as discourse, and may travel, be fished out of a garbage can and / or may be a palimpsest.

We have also noted at various points in this article that policy is assumed to flow unproblematically into implementation. Indeed, there is often a faith in an almost alchemical transformation of base policy into the pure gold of development impact simply because the words of the policy have been invoked. Yet, again, the world of academic accounts of policy, implementation and change are full of notions of policy as rhetoric; of implementation as resistance and reinterpretation; and of change as messy and reversible.

These accounts of how policy is made are those that we are grounded in as academics but they have the potential for leading us down a path towards the inevitable incommensurability of academic and policy knowledge. It can encourage international and comparative educators to see their proper role as theory-based critics of policies that must always fail due to their unsafe epistemological groundings; whilst policymakers and officials will tend to take an understandable counter view that this research is neither rigorous nor relevant. This, of course, would mirror much of the debate about educational research in England that began in the late 1990s (e.g., Hargreaves, 1996; Hammersley, 2002; Whitty, 2006). However, as Whitty argued in his inaugural presidential address to the British Educational Research Association, it is important to try to break out of the impasse between entrenched positions stressing the fundamental divide between policy and academia. This is the subject of our concluding comments.

CONCLUDING THOUGHTS: THERE MUST BE SOME KIND OF WAY OUT OF HERE

The SADC/UNESCO project that has been the principal source for our reflections in this article sought, albeit largely unconsciously, to reject the divide between academic and policy knowledge. The project was on the face of it delivered through a consultancy mode that could easily have privileged only policy knowledge. However, there was a desire to engage with academic knowledge and this led to a conscious decision to choose a largely academic team of regional consultants to lead the project, and to include a review of research literature as a key method for data collection. Moreover, throughout the process, there was a clear sense that the project team, and particularly the project director, were seen as having a powerful knowledge resource that merited respect even where this was not the kind of knowledge that the commissioning agencies themselves saw as the most valuable.

For us as the lead consultants too the process was one of repeated bridging and translation work, across the boundaries of what we considered robust data and knowledge; the attempt to meet consultancy norms regarding speed, cost and client relationship maintenance; and our commitment to a genuine dialogue with policy and policymakers.

Together as commissioners of this work and as researchers we constructed a methodology that sought to balance political and scientific imperatives; to balance local contexts and regional comparability; and to use multiple sources of data and approaches to analysis and reporting in order to provide a better basis for policy. We embraced the dualism that Raffe saw as being at the heart of policy learning and tried to make the tension a creative one.

In reflecting on the project, we still believe that this was successful and worthwhile.

Nonetheless, the project has also served to stimulate us to reflect more widely on the policy-research relationship. Like Whitty, we reject the attempt to recreate research in the policymaker's likeness. We cannot but see research and policy through critical lenses such as power, but we also cannot walk away from the challenge of seeking to help make better policy.

Particularly for VET in Africa, where our case study was located, the scope for "what works" or statistically-driven approaches to policy and practice is almost non-existent at present. This may suggest that those development cooperation agencies that are increasingly wedded to such approaches need to realise that rather than criticising researchers for failing in rigour, they should make a serious attempt to fund the kind of research that they think matters or be more realistic in their notion of useful knowledge. They do need to consider approaches to such research that strengthen, national policymaking capacities.

For policymakers, there is a similar challenge in investing in data, something that the SADC Ministers have acknowledged and UNESCO has recommended at the global level. However, there is a further challenge of thinking more critically about what the relationship between policies and impacts are and being prepared to explore this even when it is uncomfortable.

However, we want to finish with a challenge to our own community and to reiterate that there are pressing needs for international and comparative educators to be clearer in seeking to shape research agendas; more rigorous in their approaches to research; and better in their external communication of their findings. As this is a special issue on VET, it is important to reiterate the particular challenge of reawakening research on VET-for development from twenty years of slumbers.

ACKNOWLEDGEMENTS

REFERENCES

Akoojee, S., 2012. This issue.

Akoojee, S., Gewer, A., McGrath, S.A., 2005. Vocational education and training in Southern Africa: A comparative study. Human Sciences Research Council, Pretoria.

Allais, S., 2010 The implementation and impact of National Qualifications Frameworks: Report of a study in 16 countries. ILO, Geneva.

Ball, S., 2003. The teacher's soul and the terrors of performativity. Journal of Education Policy 18, 215–228.

Billetoft, J., 2011. TVET reform in Sub-Saharan Africa: the case of Mozambique. Paper presented to the UKFIET Conference on International Education and Development, Oxford, September.

Carton, M., 2011. Research on technical and professional skills development in Africa: from decline to new perspectives. Paper presented to the UKFIET Conference on International Education and Development, Oxford, September.

Chakroun, B. (2008) What can we learn from policy learning? In ETF Yearbook 2008. Policy Learning in Action. ETF, Turin.

Chakroun, B., 2010.National Qualification Frameworks: from policy borrowing to policy learning. European Journal of Education 45, 2, 199-216.

Chisholm, L., Leyendecker, R., 2008. Curriculum reform in post-1990s sub-Saharan Africa. International Journal of Educational Development 28, 195-205.

Connell, R., 2007 Southern Theory. Polity, Cambridge.

Crossley, M., Broadfoot, P., Schweisfurth, M., 2007. Changing educational contexts, issues and identities: 40 years of comparative education. Taylor and Francis, London.

Crossley, M., Jarvis, P., 2001. Context matters. Comparative Education 37, 405–408.

Dale, R., 1999. Specifying globalisation effects on national policy: a focus on the mechanisms. Journal of Education Policy 14, 1, 1-17.

Dunkel, T. 2009. Reviewing European VET Policy – challenges for comparative analyses. In P. Grollmann and M. Hoppe 2009. Methods and Instruments for the evaluation and monitoring of VET systems. Conference Proceedings. Königswinter, Germany 10-11 December 2009. Federal Institute for Vocational Education and Training, Bonn.

Glass, G., 1987. What Works: politics and research. Educational Researcher 16, 3, 5-11.

Goldstein, H., 2008. Evidence and education policy – some reflections and allegations. Cambridge Journal of Education. 38, 3, 393–400.

Gorard, S., 2008. Who is missing from higher education? Cambridge Journal of Education, 38, 3, 421-437.

Grootings, P., 2004. Learning Matters ETF Yearbook 2004. ETF, Turin.

Hammersley, M., 2002. Educational Research, Policymaking and Practice. Paul Chapman, London.

Hargreaves, D., 1996. Teaching as a Research-based Profession: Possibilities and Prospects. Teacher Training Agency, London.

Hippach-Schneider, U. 2009. Challenges and Chances of a European monitoring instrument in VET. In P. Grollmann and M. Hoppe 2009. Methods and Instruments for the evaluation and monitoring of VET systems. Conference Proceedings. Königswinter, Germany 10-11 December 2009. Federal Institute for Vocational Education and Training, Bonn.

Jansen, J. 2005. Targeting education: the politics of performance and the prospects of Education for All. International Journal of Educational Development 25, 4, 368-380.

Lewin, K., 2011. Policy dialogue and target setting: do current indicators of Education for All signify progress? Journal of Education Policy 26, 571-587.

Lim, Z. 2012. This issue.

McGrath, S., 2010a. Beyond aid effectiveness: The development of the South African further education and training college sector, 1994–2009. International Journal of Educational Development 30, 525-534.

McGrath, S., 2010b. The role of education in development: an educationalist's response to some recent work in development economics. Comparative Education 46, 237–253.

McGrath, S., 2011. Where to now for vocational education and training in Africa? International Journal of Training Research 9, 1&2, 35-48.

McGrath, S., 2012. This issue.

McGrath, S., Mason, M., 2011. The pragmatics of education journals: the case of the International Journal of Educational Development. Southern African Review of Education 16, 2, 45-60.

Norrag, 2012. Norrag News 46: Towards a new Global World of Skills Development. http://www.norrag.org/news/37/nn46-towards-a-new-global-world-of-skillsdevelopment-now-online_.html

Ozga, J., 2009. Governing education through data in England: from regulation to self-evaluation. Journal of Education Policy 24, 149-162.

Powell, L., 2012. This issue.

Raffe, D., 2011. Are "communication frameworks" more successful? Policy learning from the Scottish Credit and Qualifications Framework. Journal of Education and Work. 24:34, 283-302.

Raffe, D., Spours, K., (Eds.) 2007. Policy-making and policy learning in 14-19 education. Bedford Way Papers, Institute of Education, London.

Rose, N. 1991. Governing by numbers: figuring out democracy. Accounting organizations and society 16, 7: 673–92. 21

SADC, 1997. SADC Protocol on Education and Training. SADC, Gaborone.

SADC, 2005. Towards a Southern African Development Community Qualifications Frameworks: Concept Paper and Implementation Plan. SADC, Gaborone.

SADC, 2007. Towards a Framework for the improvement of Quality Assurance Systems in Member States. SADC, Gaborone.

SADC/UNESCO, 2009. Terms of Reference to conduct an assessment and review of Technical and Vocational Education and Training (TVET) in the SADC Region and development of a regional strategy for the revitalisation of TVET. Downloaded November 2010 from http://www.sadc.int/

Schoenfeld, A., 2006. What doesn't work: the challenge and failure of the What Works Clearinghouse to conduct meaningful reviews of studies of Mathematics curricula. Educational Researcher. 35, 2, 13–21.

Schweisfurth, M., 2011. Learner-centred education in developing country contexts: From solution to problem? International Journal of Educational Development 31, 5, 425-432.

Shoesmith, K., 2011. Towards a taxonomy of skills. Paper presented to the UKFIET Conference on International Education and Development, Oxford, September.

Speer, S. 2009. Peer Evaluation: Facilitating Changes on the System Level. In P. Grollmann and M. Hoppe 2009. Methods and Instruments for the evaluation and monitoring of VET systems. Conference Proceedings. Königswinter, Germany 10-11 December 2009. Federal Institute for Vocational Education and Training, Bonn.

Thomas, G., Pring, R. (Eds.), 2004. Evidence-Based Practice in Education. Open University Press, Maidenhead.

UNESCO, 2012. TITLE TO BE FINALISED

UNESCO-BREDA, 2010. Regional Contribution to Statistical Information Systems for Technical and Vocational Education and Training. UNESCO-BREDA, Dakar.

UNESCO-UNEVOC, 2011. Nairobi Pledge for Action. The role of UNEVOC Centres in furthering the Second Decade of Education for Africa TVET Plan of Action, Nairobi – Kenya, 13-16 December 2011. Downloaded 29 December 2011 from http://www.unevoc.unesco.org/fileadmin/user_upload/docs/Final_Draft_Nairobi_Commit ment_to_action_approved.pdf

Whitty, G., 2006. Education(al) research and education policy making: is conflict inevitable? British Educational Research Journal 32, 159-176. 22