John Killeen Commemorative Lecture 2005

‘Known knowns and known unknowns’: What can evaluation tell us about labour market impact?

Jim Hillage

It is with some trepidation that I accepted the invitation to give this lecture. I did not know John Killeen, but I do know of, and have always been impressed by, his work and we sought his advice on the recent Institute for Employment Studies (IES) study on the impact of advice and guidance, but by then he was unfortunately too ill to help.

My trepidation increased when, in preparing for this lecture, I realised:

a) how little I knew about careers information, advice and guidance, although as a jobbing researcher and evaluator I have been involved with a number of studies in the area over the years, and

b) how much John had contributed to our understanding in this area.

By then, however, I had already chosen as my text the now famous quote by Donald Rumsfeld, US Defense Secretary:

‘Reports that say that something hasn’t happened are always interesting to me because we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns - the ones we don’t know we don’t know.’

This quote, rather unfairly I think, won the Plain English Campaign’s ‘Foot in Mouth’ Award in 2003. Unfair because I think it is actually clear and certainly less ridiculous than Chris Patten, ex-Conservative Cabinet Minister and Governor of Hong Kong, who was runner up with:

‘The Conservative Party has committed political suicide and is now living to regret it.’

Now I am not suggesting that there is any similarity between John Killeen and Donald Rumsfeld. However, I like to think he would have agreed that with the idea that a researcher’s job is to understand what we know and concentrate on the things we don’t know and identify the things we did not realise that we needed to know. After all John finished one of his last published works with the statement:

‘More research required is a common conclusion, but it is important that this does not keep starting at square one - of this we have had enough.’

(Hughes et al., 2002)

So I set myself the task of reviewing what was known and not known about the labour market impact of information, advice and guidance (IAG). With the Government’s upcoming review of adult guidance and the recent publication of the baseline report on IES’s study on measuring the impact of advice and guidance it seemed an appropriate, if not original, brief as there have been a few reviews over the years. However, I will present my interpretation of what we know or don’t know not just in terms of the evidence, including some of the more recently published studies and other relevant research, but also in terms of the research techniques which underlie it.

I thought I would start with a model.

At the heart of any evaluation is a model which seeks to map the link between the intervention, in this case careers information, advice and guidance, and what happens as a result, in both the short-term and the long-term.

In any evaluation the model is determined by the nature of the intervention and the reasons for making it – its aims and objectives, i.e. what you are trying to achieve and how it is to be measured.

It still surprises me in this age of evaluation how difficult policy-makers can find it to answer this question clearly (and even more the practitioners who are tasked with implementing the policy). This probably reflects the multiplicity of policy goals that the designers hope to achieve. It does not help that, when they are expressed, the goals are generally set at a high level, e.g. to improve productivity or social inclusion, while the objectives are not clearly linked to either the intervention or the goals and/or are subject to change as the policy framework develops. While unspecified outcomes certainly make the job of the evaluator more difficult, I’m not sure they help the people on the ground trying to deliver the policy either. A clear ‘line of sight’ between strategy, policy and practice serves to explain why certain actions are being undertaken or certain groups are being targeted and, therefore, improves the likelihood of effective implementation.
I refer to the model for three reasons:

- Firstly, it helps to marshal the evidence across the chain of impact and to identify the known knowns and, therefore, the unknowns.
- Secondly, it illustrates the problems faced by those tasked with assessing the impact of information, advice and guidance.
- Thirdly, it may help to identify possible ways forward.

So as a researcher we are interested to see who takes part in a given intervention, their immediate reactions, what skills they gain or attitudinal change place as a result leading to changes in behaviour such as looking for a job or taking up a training course. In the literature, the precise components and titles of each of the links in the chain – the individual boxes – vary and although this serves as a general model, more specific indicators would apply to the evaluation of specific interventions.

There are obvious problems with evaluations across the chain at this level, e.g.:

- the time lag between intervention and behavioural change and subsequent labour market activity. A person may as a result of a careers intervention decide to change their job but it may not happen immediately, although you would hope to be able to see some intermediate changes, e.g. in their job search behaviour, let alone...
workplaces to compare what the HR Department says is the policy or practice on a particular issue and the views of the line managers or employees on the ground. The differences can often highlight the gap between the HR rhetoric and the workplace reality.

But the problem of defining the intervention is not just one of language or administrative records. The nature of the intervention varies in so many ways, for example by:

- **Context** – IAG is provided at work; in educational settings; in the community. It is also often provided in a variety of circumstances, e.g. as part of programme support to the unemployed or linked to some form of learning provision. This gives us a problem of isolating the IAG intervention.
- **Nature of the provider** – ranging from employers (where IAG is generally delivered by a line manager), training providers, specialist IAG providers, Job Centres, friends and colleagues, etc.
- **The type and intensity of the provision** – IAG covers a wide range of support, from simple signposting or the provision of information, through specific help with CV preparation or identifying learning or employment opportunities through to a series of investigatory and guidance sessions combining a range of activities.
- **Quality of the provision** – and the knowledge and skills of the provider.
- **Form of provided** – face-to-face, telephone, computer, written materials, etc.
- **Frequency of provision** – e.g. one-off or time-limited provision or a series of interventions.

The list is almost endless and compounded if IAG is viewed as a process rather than a discrete activity, i.e. a series of interventions and personal reflection over a period of time. That said, most significant labour market interventions (e.g. training or job search assistance) could be seen more as a process than a single activity. However, the more the intervention is boundless and multi-faceted, the more difficult it is to define and the more difficult the impact is to measure.

Therefore, from a research point of view, the greater the homogeneity of the intervention (and the larger the consensus between provider and participant about what actually happened) the better. The wider the range of provision being investigated the greater the difficulties for the researcher as the intervention then becomes a range of variables rather than a single variable in any explicit or implicit equation. In the ETP evaluation we faced a number of problems in trying to identify the effect of the IAG element as provision varied dramatically across the pilots. Some provided step to learners either face-to-face or in the form of written materials at the start of training. In other areas there was a ‘call-down’ service with learners given a phone number to call if they wanted support (and few did). Some only provided information or advice and not guidance because they thought it would put employers off. Others attempted to give all learners access to an IAG provider at the end of their training, although in some areas this was part of an ‘exit’ interview than a ‘next steps’ interview.

I draw two lessons here:

- do not assume that a programme will be universally applied, especially when there are different providers supplying a notionally similar service.
- tracing the impact of IAG as a general process or what a range of individuals receive is far more difficult than tracing the impact of a specific IAG measure targeted at a specific group of people.

Again John provided us with indications of the way through this problem by focusing sharply on one particular intervention and collecting a wide range of data around it. However, I am not sure I would agree with this, it also highlights to me the importance of qualitative research in this respect.

**Who takes part?**

Having established the intervention we are obviously interested in who takes part, particularly compared with who does not and ultimately the effect the intervention has on them.

In passing it is interesting to note that according to the recent MORI survey (Taylor et al. 2005), IAG users are more likely to be: young, relatively well-qualified, have recently taken part in training or learning and more likely to have higher aspirations in terms of their future career. They already have a reasonable stock of human capital and are either looking to acquire more or looking for ways to exploit better what they have. This would seem to bear out the ‘wise search’ hypothesis that John Killeen articulated in the Gateways evaluation (Killeen, 1996), i.e. that people who use guidance provided under any voluntary programme may be demonstrating the superiority of the general search strategy. There is less evidence that guidance (or at least information, advice and guidance in the MORI survey) is used by John’s other ‘deficit’ groups, i.e. people who need a remedial activity to help cope with their adverse labour market circumstances.

This begs a question for me about whether guidance is reaching the people who need it most and what interventions work best for the ‘hard-to-reach’.

**Outcomes**

Moving on to the outcomes of guidance, the key question in measuring impact, clearly articulated by John Killeen and Michael White in their study of the impact of guidance on employed adults, is what would have happened in the absence of the intervention – i.e. the counterfactual position (Killeen and White, 2002). The new established way of measuring this is through a comparison between the participants and a control group. The trick is to ensure that the controls mirror the participants.

The problem is that establishing a control group from the general population builds in an immediate problem of selection bias. Voluntary participation in IAG is often triggered by either the desire or need to change and, in that respect, participants are qualitatively different to the rest of the population.

The gold standard is to establish an experimental design and randomly assign individuals to the participant group or the control. The control group provides the counterfactual data on what happens in the absence of the intervention and any subsequent differences in outcomes between the two groups can, therefore, be ascribed to the intervention. Such evaluations are common in the US for adult education programmes, but very rare in the UK, primarily due to practical and ethical considerations and the speed of the policy development/implementation process.

There are ways round this problem, e.g. by operating policies in certain institutions or areas of the country and not others, rather than organising assessment at the level of the individual.

For instance the Employer Training Pilots were established in their second phase in a number of LSC areas which were specifically chosen for their similarities to two control areas we had selected. Data were collected before the pilots started in the pilot areas and the control areas and then one year later to see whether there was any relative difference in the established effect measures (e.g. the provision of qualification-based training to employees without Level 2 qualifications).

These sort of methods could be adopted to test any new policy developments on IAG but they do involve designing the policy with the evaluation in mind. All too often policies are pulled in after a policy has been designed or even started and therefore have to start after the event or a policy is applied in a such a way that it is impossible to establish effective controls or baselines.

In the IES impact study, we had no such luxury of an experimental design. The design does attempt to control for selection effects, by comparing information users with advice and guidance users.

There may still be differences between the two groups, although we have tried to control for that by matching individuals in the two groups through a propensity score matching process and pair off individuals in both groups with similar characteristics (e.g. age, gender, employment and learning experience, etc.). However there may still remain unobserved differences between the two groups which may prove important in explaining any differential outcomes.

One problem we had was that the baseline data collection took place after the intervention, in order to allocate respondents either the information or the advice and guidance group. In the matching process we decided not to control for various attitudinal measures which could have been affected by the intervention itself. This has led to some criticism that we have not matched like with like and the ‘treatment’ group of advice and guidance recipients could be qualitatively different than the control, e.g. they were generally more positive and this could explain some of the initial findings.

However, if further funding is forthcoming for the planned subsequent waves we can re-analyse the data to assess the impact of and try to control for this issue. It is important to remember that we have only drawn a baseline so far, and the initial findings are not based on tracking participants over time.

We don’t yet know whether there will be a further wave of the study, although there are some indications that there might be. In longitudinal studies, as this is designed to be, it is important to minimise sample attrition which can be significant and I am pleased to say that we have had just indications that the DfES is going to fund us to send a re-contact card to the survey participants.

**What do we know?**

Do participants like what they get?

Lots of surveys report high levels of satisfaction among users of IAG services and users generally find the service they receive helpful or useful (Taylor et al. 2005, Milburn et al. 2003, Barnett, 2005).

In the early 1990s I was involved in a study about developing approaches and tools for measuring satisfaction with the then Careers Service. We encountered some concern among the careers community. In particular it was argued that a good guidance session could be quite challenging for the client, especially if they came out realising that their proposed career path of being a pilot or a rock musician was unrealistic – the so-called ‘positive negative’ – and they might register negative satisfaction, although the session may be seen in the long-term as useful. I note that most measures now talk about how useful or helpful the service was and I also note that they generally find very positive responses.
Obviously high levels of satisfaction, however measured, are better than low ones, but it is important to set any findings in some sort of comparable context. In the Employer Training Pilots evaluation we also found high levels of satisfaction among the ETP learners, a fact that has been highlighted by Government as indicating the value of the initiative. However, the satisfaction levels are no higher than that recorded (using a similar scale) by the LSC surveys of learners in FE colleges and work-based learning.

Participants in development activities of any sort generally report that it was a positive experience. A further note of caution into reading too much into such immediate measures comes from the training evaluation literature which suggests that there is relatively little correlation between learner reactions and measures of learning or subsequent measures of changed behaviour.

What does it mean? Do we know what drives satisfaction?

However it can be important to see what users find most useful about a careers intervention both to contribute to practice development and identify the sources of any impact. This is an area where rigorous developmental research, which can cope with the range of variables involved, can provide particular insights. The IER longitudinal case studies (Barnes, 2005) looks like being a really useful study in this respect.

For example, it found that all but one of the clients studied thought the guidance useful with the main ingredients being:

- Exploring and challenging client perceptions, together with increasing awareness of learning or employment opportunities;
- Giving clients access to networks, information and knowledge enabling them to be better informed;
- Encouraging constructive change, e.g. increasing the client's self-confidence; developing skills; developing understanding which broadened ideas; as well as motivating, inspiring and encouraging the client;
- Providing the client with a positive experience by: creating the opportunity for reflection and in-depth discussion; and by reassuring, confirming and/or clarifying plans and/or progress.

The key appeared to be the skills of the practitioner.

This finding is also reflected in the MORI survey (Taylor et al. 2005) which found that around seven in ten IAG users of found it helped them decide what to do next and increased their awareness of training and job opportunities. Users of more formal IAG services were significantly more likely to find the service helpful than those who had used more informal sources. In a regression analysis, some of the key drivers in generating positive outcomes appeared to relate to the quality of the service offered, i.e. the knowledge and professionalism of the provider, the provision of a comprehensive and accessible service.

Another interesting point to emerge from their analysis was that participants with the lowest level of qualifications were much less likely than others to have found the service they had used helpful. Unfortunately, it is not clear why, but as this group could, at least in theory, benefit from any support IAG provided in getting them started on the learning ladder it would be interesting to know more about what works with these individuals.

In the evaluation of the New Deal for Young People, Van Reenen (2001) was able to quantify the effect that the job assistance element had on the flow into employment and show that it was more important than other elements of the New Deal such as the job subsidy. They were able to make the finding partly because it was a well-funded study that looked at a range of factors but also because they were able to isolate the specific intervention and focus on clear effect measures.

Although, Killeen and White (2002) found that guidance participants benefited through increased entry rate into education and training, producing an enhanced rate of qualification, we appear to know less about the role of guidance on attainment. The reviews I have read suggest that 'hard evidence is limited here' (Maguire, 2004). If this is right, then it seems surprising and a link in the chain on which it is worth concentrating.

The IES impact study although the advice and guidance group felt better able to make a change to their job or learning, they were no more likely to have made progress than the information group. For example, in each case about a half had started a training course. However, when asked, the advice and guidance group were significantly more likely to attribute their participation in a course to the intervention they had received, particularly younger people and those with a recent history. This suggests two things:

- the information group were more likely to have done the course anyway, without the help they had received - they were already set on what they intended to do, or closer to doing it and the support they received, therefore, made less of an impact on their intentions;
- making a change, like taking a course or changing job, is a bigger step than to take for older people and those without a recent learning experience than younger people with recent learning experience.

Hopefully in the future we will be able to examine these and other issues like qualification attainment in more detail.

Economic effects on employment and earnings

Towards the end of the chain the evidence trail peter vacates. There is very little conclusive evidence on the longer-term economic effects of guidance. For example, Killeen and White (2002) found no effect on job satisfaction and 'no indication that the guidance group improved its earnings more than the comparison group, despite some exhaustive examination'.

Does this mean that there are no economic effects? I am reminded of the old researchers' adage that the absence of evidence is not evidence of absence.

Does it mean that it is impossible task to measure the effects so far down the chain? Well it is certainly difficult, but nothing is impossible.

For example we can again look to the US for inspiration as an evaluation of an intensive welfare to work programme in California in the 1990s was able to demonstrate, albeit through random assignment, that the programme produced a 12 per cent increase in employment and a 22 per cent increase in earnings between the treatment and the control group (Riccio et al., 1994).

Known knowns

So to sum up, what do we know? Beginning on the research front:

- Measuring the impact of guidance is difficult.
  Various reviews highlight a large number of studies, producing lots of associative findings, but few ‘killer facts’, especially further down the chain of impact. This is due to the variety of the interventions, the diversity of the participants and the complexity of the chain of impact with which available research techniques have so far been unable to cope.

- The search must go on. There is general consensus on the need to continue to look for impact and to understand the causalities involved in the field of information, advice and guidance. It is important for policy justification and allocation of public expenditure, with the forthcoming comprehensive spending review and the current commitment to evidence-based policy in mind. It is important for the design of policy, with the review of adult guidance in mind. If more had been known about how best to provide IAG in the workplace, what the demand was and how best to meet it, the IAG element in the Employer Training Pilots would have been more effective. It is also important for and will help focus the actions of practitioners on what works best to meet specific policy goals.

In so doing I think there is also a fairly strong consensus on the ways forward:

- more longitudinal research and tracking studies, and hopefully further waves of the IES impact study and the IER longitudinal case studies can contribute in this respect. The problem is that such studies take time (and money) to generate substantive findings - policymakers can’t wait. But they are still worth doing as the same policy questions keep coming back.

- more focused studies concentrating on specific groups and specific interventions and on specific outcomes, trying to trace a series of thin lines of impact along the general chain or breaking up the chain into manageable chunks. A number of things would help here:
the establishment of clearer terminology and typologies about the types of intervention made
the development of standardised measurement criteria to clearly assess whether progress has been made so we can better aggregate findings and set them in a comparative context
improving administrative data for example on who has received what intervention and linking that data to other labour market actions the participant may take – the single learner record will be useful here.

- Linking in with other areas of research, which has already established the links further down the chain. For example, there is now a considerable body of knowledge about the rate of return on earnings of obtaining specific qualifications. If clear connections between the receipt of guidance and qualification attainment at specific levels can be established, we can use existing knowledge to infer the wider impact.
- The process does not just require quantitative research. There is a continuing role for well-conducted smaller scale studies and qualitative research, e.g. to investigate what people actually get from a guidance intervention or to understand more about the process of impact and movement along the chain. Such studies are now especially relevant with the development of meta-analyses and systematic reviews which can be used to collate findings and make them more generalisable. We are currently conducting a systematic review at IES and it is a tortuous process but it has highlighted the value of tight research questions and clearly articulated methodologies if you want to aggregate research findings – putting pieces of the jigsaw together to obtain a clearer picture.
- Better research also needs the help of policymakers not just in the involvement of evaluators in policy design but also the establishment of clearer intermediate outcome measures, starting with what you are trying to achieve and then designing an intervention to fit the target. Fewer, but larger, initiatives with time to make an impact would help too. When you are trying to see the wood for the trees the last thing you want is people planting more trees!

But as John reminded us in conducting further research we should build on what we know and concentrate on what we don’t know. In the realms of guidance we know that:

- IAG can be effective when combined with other interventions – but how does this work and why and in what circumstances?
- Where an effect can be measured, bigger interventions such as guidance have more of an effect than just information - but there is more scope for finding out about different techniques and what works best and why and in what circumstances.
- Quality of provision is important – but how to secure it and what is the effect of quality standards?

**Known unknowns**

And there are many things we still know that we don’t know much about, and I highlight just three:

- What happens in the workplace – important for policy as it focuses more workplace interventions, but also for employers. Workplace interventions are even more focused on the wise searchers, for example the well-qualified. There may be large amounts of unidentifiable need as well as unmet need among the less qualified and those under threat of enforced job change.

- What works best for people most disadvantaged in the labour market – particularly older workers at risk of enforced job change or individuals without any qualifications? How can IAG reach the hard-to-reach and where it finds them, can it help them stock up or restock their human capital in such a way as to equip them better for the future?

- Cost and cost effectiveness – although some forms of IAG may be more effective than others, taking costs into consideration may produce a different conclusion. Although the GAIN program in California I referred to earlier found positive employment and earnings effects, the net cost benefits (i.e. taking into account the costs of the programme and the reduction in welfare payments, etc. to participants) were only positive in two of the six areas where the programme was introduced. Examples in the UK of cost-effectiveness are particularly rare.

**By definition, I don’t know what the unknown unknowns are!**

There is no magic bullet. No one study will provide the map in our journey towards understanding:

‘The general value of guidance, as of other programmes, can only be established by repeated studies. A single study contributes to the build-up of evidence, and may stimulate further investigations.’

(Killean and White, 2002)

As I have tried to show, the terrain is just too difficult to map out the single path in one attempt. The destination is slightly different with each study (though generally in the same broad direction). It is impossible to trace general impact with a single study – but studies can and should build on each other and add stepping stones towards the goal. However, we have made progress and it is important that we regularly take stock, reflect clearly on what we know (and don’t know), concentrate on the latter and move on. There are some known knowns, some stepping stones towards the goal of understanding the impact of information, advice and guidance.

John laid more than his fair share of these stones. Let’s stand on them and move forward.

**References**


**Notes**

John Killean was a senior fellow at NIEEC from 1977 to 2003. The next commemorative lecture will be given in October 2006 by Jenny Kidd who will collaborate with John on numerous projects. Details will be available on the CRAG/NIEEC website later this year.

**For correspondence**

Jim Hillage
Institute for Employment Studies
Mantell Building
University of Sussex
Brighton
BNI 9RF
Tel: 01273 686751