Editorial:
Unlearning Some of our Social Scientist Habits

E. Jane Davidson
Davidson Consulting Ltd., Aotearoa/New Zealand

According to the organizational learning literature, some of the most important learning an organization can do is in the form of organizational *unlearning*—getting rid of erroneous or dysfunctional knowledge, habits, and practices. Simply adding new skills, behaviors, and ways of thinking isn't enough. Without some explicit attention to the process of unlearning, existing knowledge and know-how becomes a barrier to constructive change and new learning.

The same is true in evaluation. Many of us had our primary training in the social sciences, which imbued us with a set of knowledge, skills, practices, and habits that are considered good practice in those disciplines. When we moved into evaluation, we became aware that there was something more that we needed to know (apart from social science research skills), and many of us set about adding to our repertoire so we could practice good evaluation. But how well have we paid attention to *unlearning* some of the baggage we brought with us from the social sciences?

Four social scientist habits that I often see as barriers to good evaluation are:

1. Including models or theories but not using them evaluatively
2. Leaping to measurement too quickly
3. Reporting results separately by data type or source
4. Ordering evaluation report sections like a Master's thesis

1. Including Models or Theories but not Using Them Evaluatively

Our social science training teaches us that there's nothing as useful as a good theory. Therefore, a common feature of evaluation reports is some sort of literature review and often the presentation of a theoretical model that illustrates how some of the variables fit together.

While I am all for having a conceptual framework to help explain a piece of work—and I often use them myself—in my experience, many of these theoretical models (especially if they are neither representations of program theory nor of the wider system in which the program is embedded) are often only tangentially related to the evaluation questions; that is, what the client really needs to know. As such, these theoretical models have a tendency to gravitate evaluation effort away from the 'core' task of answering important evaluative questions and into the territory of "wouldn't it be nice to know" research questions (e.g., about relationships among variables).

In my experience, the worst examples of low utility models and theories tend to turn up in cases where a university faculty member has been co-opted onto the evaluation team for their *content* (but not evaluation!) expertise. It is often clearly evident that the primary purpose of including the model is to allow said faculty member to measure several variables in the

model and publish a paper on the results, rather than to enhance the validity, utility, or comprehensibility of the evaluation.

So, what would constitute good use of models and theories in evaluation? The most useful examples I have seen are representations of program theory (whether linear or more systems-oriented) that (a) succinctly clarify what the evaluand is and how it is expected to generate outcomes of value and (b) are linked directly with important evaluative questions; that is, questions about the quality or value of the evaluand and/or its effects.

Too often, logic models and other theories are presented in an evaluation but left dangling in the introduction section and not used later on. Or, they can be used inappropriately, for example, when too much effort is spent evaluating the theory at the expense of the main task—evaluating the program. If the model is unrelated to the main evaluation questions, it can actually get in the way of a concise and useful explanation by sending the reader off on theoretical tangents.

2. Leaping to Measurement too Quickly

Perhaps this has to do with the literature review → methods transition in traditional applied research, but I have been staggered at how often I see evaluations leap directly to measures without even the slightest attention to evaluation questions. It's almost as if the instructions have been simply "Evaluate this" and the first step has been identifying indicators or variables of interest and figuring out how to measure them. Or, in the case of qualitative research, identifying key informants and writing up lists of things to ask them.

The problems with the "leap to measurement" approach quickly become apparent when we reach data collection and analysis. The write-up consists of either screeds of means and standard deviations, perhaps some correlation matrices, and/or oceans of stories or quotes. But the interpretation step from here to "So what? What does this tell us about the value of the program?" is completely missing. Perhaps the writers feel that the program's merit should be intuitively obvious, or that everyone should be allowed to make their own interpretation. But for the client seeking to make evidence-informed decisions using the findings, it's not clear what questions the measures were designed to answer in the first place, let alone what answers they might provide.

Evaluations cannot produce useful answers unless they actually ask useful questions in the first place!

Sometimes we see a list of [descriptive and/or causal] research questions rather than truly evaluative questions. These are often asked at the variable level (e.g., "What is the effect of Program X on Variable Y?"), in contrast with evaluative questions about the program (or other evaluand) itself (e.g., "To what extent does Program X produce outcomes of sufficient value to justify its cost?"). Questions about the nature or relationships among variables are what publishable research is all about. Questions about the quality of the program, the value of the outcomes, its comparative cost-effectiveness, key learnings, and the areas for improvement are what evaluation is all about.

So, a crucial but often-skipped step on the way to measurement (or, more correctly, qualitative and quantitative data gathering) is the formulation of a set of evaluation questions to guide the project. These should be developed in (or at least after) consultation with the primary intended users of the evaluation. Each question should be something that an actual decision maker needs an answer to. And each question should be evaluative (not merely descriptive) in nature—these are the questions that are most useful and relevant for decision making.

It is true that good research training will teach people how to write clear research questions before leaping to measurement; the key 'unlearning' task here is letting go of the tendency to write questions about variables and

the relationships among them. What an evaluation needs are questions about the quality of the program, the value of its outcomes, its comparative cost-effectiveness, its sustainability, and so forth.

3. Reporting Results Separately by Data Type or Source

In the applied social sciences, as in evaluation, we are all taught the importance of 'triangulation'—using data of different types and from different sources. But what we are not usually taught is how to weave the findings together to create a cohesive answer to a real question. Time and again I see results reported separately by data type and source, e.g., in sections entitled *Qualitative Findings* and *Quantitative Findings* (separating types of data) or *Responses from the recipient survey* and *Responses from the instructor survey* (separating sources of data), with no attempt to link them together.

The whole point of triangulation is to get different perspectives on the answer to the same question. It follows that, in order for triangulation to serve its purpose, different types and sources of data *must* be woven together in the analysis. The end result should be a cohesive answer to an evaluation question.

It strikes me that it is not so much a matter of unlearning what we were taught about triangulation in social science research classes, but acquiring a skill that we should have been taught in order to do good descriptive research as well. The evaluative aspect (asking explicitly about quality or value) will be different, of course, but making effective use of triangulated data should be a fundamental skill taught in even the most basic of research courses.

4. Ordering Evaluation Report Sections Like a Master's Thesis

It has always seemed somewhat bizarre to me that the majority of Master's (or Honours[1]) students in the social sciences plan to graduate and find work in an applied setting, yet graduate training at the Master's level seems so strongly geared to preparing people for a doctorate or a career in academic-style research. A lot of effort goes into teaching people how to write scientifically using sophisticated-sounding language, conforming strictly to the almighty American Psychological Association Publication Manual, and generating publishable papers. But how often are Master's students taught to write a report that gets straight to the point and will make sense to managers with no social science training?

This was the piece of social science research unlearning that came to bite me swiftly and mercilessly when I first moved into an applied setting to use my new-found skills in applied psychology. My reports were written and structured in a way that would have delighted my professor but did nothing but irritate management and minimize the likelihood that they would read, let alone use, my findings.

In our social science training we are taught to lead the reader carefully through the theoretical and empirical backdrop for our study, the research questions, the methodology, the results (be careful to be very neutral and not draw conclusions at this stage!), then the discussion section where we may draw tentative conclusions about what the data "appear to suggest" (but never actually say directly that something is the case, even if we are certain to a level of $p < .001$!).

I occasionally work with consultants who were similarly trained in the applied social sciences and I frequently find that the draft 'evaluation' reports they send me are plagued with the structure of a Master's thesis, which makes it incredibly difficult for me (let alone the client) to decipher results. It typically goes something like this: Executive Summary (lots of

[1] Honours is, roughly, a one-year post-graduate degree offered in countries using the British system.

introductory information, methodology, sampling, random snippets of findings that fail to give a clear sense of the program's quality or value, plus something incomprehensible about moderator variables), Introduction, Literature Review, a theoretical model and detailed explanation of the relevant social science theory explaining the links along some variables (unfortunately not a program logic model, and not even remotely linked to an evaluation question—this part contributed by a university faculty member with no evaluation expertise), Methodology, Findings (about 20 pages of raw data, all presented separately by source and data type with virtually no explanatory narrative, none of it linked back to the questions), Conclusions (some glimmers of hope in here, but by now we are 37 pages into the report and have lost most of our audience), Appendices.

For the client, reading a report like this feels like wading through mud. Page after page of graphs and interview quotes, but not a hint of whether or how they were used to answer any question of value. When, oh when, are they going to get to the point?

How can we get our evaluation reports to make more sense to clients?

One strategy I use is to structure the *Findings* part of the evaluation report into 7 +/- 2 sections, one for each of the 'big picture' evaluation questions used to frame the evaluation. In each section, all data pertaining to that question (qualitative, quantitative, interviews, surveys, observations, document analyses, from different people and perspectives) are presented, interpreted as they are presented, and woven together to form a direct answer to the question. Next, I write a 2-page executive summary using the same structure: 7 +/- 2 questions with straight-to-the-point and explicitly evaluative answers of 1-2 paragraphs each.

If the client has seven or so major questions about the program that need to be answered, then the first two pages he or she reads (perhaps the only two pages!) should contain direct answers to those questions. And if the client wants to know on what basis those conclusions were drawn, it should be a simple matter to turn to the relevant section of the report and see clearly how 'quality' and 'value' were defined for that particular question, what data were used to answer it, and how they were interpreted together, relative to those definitions of quality/value.

Obviously there are many different reporting structures that will meet the needs of clients; I am simply suggesting one that works well for me in some evaluation reports. The main point here is that the traditional APA format research paper, with its structure that forces clients to suffer through all the theory and methodology before they get anything resembling an answer to what they want to know, tends to be one of the least useful formats possible.

Training in the applied social sciences provides a wonderful starting toolkit for a career in evaluation, albeit one that needs topping up with several essentials. But effective topping up is best done after we identify a few social scientist habits that get in the way of good evaluation and do some work to weed them out (or, put them aside for more traditional research projects). With a little unlearning and the boldness to try something radically different and fundamentally client-oriented, our evaluation work could get really interesting—and useful, and credible, and more valid.