Theory & Psychology

Article

Evidence-Based Practice in the social sciences? A scale of causality, interventions, and possibilities for scientific proof

Theory & Psychology
2017, Vol. 27(5) 581–599

DOI: 10.1177/0959354317726876
journals.sagepub.com/home/tap


Agnes Tellings
Radboud University, The Netherlands

Abstract

This article discusses Evidence-Based Practice (EBP) in the social sciences. After a brief outline of the discussion, the work of William Herbert Dray (1921–2009) is examined. Dray, partly following Collingwood, worked on different forms of causality and methodology in historical explanation (in comparison to the social sciences), based on a distinction between causes and reasons. Dray's ladder of rational understanding is also explored here. Taking his argumentation further and sometimes turning it upside-down, a scale of forms of causality is developed with accompanying types of interventions and possibilities for scientific proof of their effectivity. This scale makes it possible to weigh interventions regarding the degree to which "hard" scientific proof is possible for them. The article concludes with a brief discussion of how interventions in psychology and education should be chosen and can be justified, both those that do and those that don't lend themselves to empirical research.

Keywords

causality, William Herbert Dray, evidence-based, justification of interventions, rational understanding

More and more, the demand for Evidence-Based Practice (EBP) in the field of psychiatry, psychology, and education wins ground, especially in times when economic crisis and an aging population decrease budgets. Whether or not, or to what degree and how, one should use EBP in these and related fields is an issue with important practical consequences but also, moreover, with interesting underlying issues concerning the relation between scientific research and professional practice and the philosophical and epistemological

Corresponding author:
Agnes Tellings, Behavioural Science Institute, Radboud University, Postbus 9104, 6500 HE Nijmegen, The Netherlands.
Email: a.tellings@ru.nl

foundations of both. In the present article, based on the work of William Herbert Dray (who died in 2009), I will develop a scale of forms of causality with accompanying types of interventions and possibilities for scientific proof of their effectivity. Since interventions of which we might wish to measure the effects are usually based on a particular understanding of the "cause" of the problem for which an intervention is sought, and since empirical evidence is based on cause–consequence assumptions, looking at explanation and causality might help the discussion further.

I will first briefly outline the contrasting views in the discussion, more particularly the arguments against EBP since they are expressed more often and more elaborately than the arguments in favor. Different related issues are at stake, some of which are more relevant to my argument than others. Throughout this article, I will use the term "intervention" broadly, thereby generally referring to treatments for adults and children in the fields mentioned above.

EBP: A brief overview of the discussion

EBP is based on economic, ethical, and practical arguments, sometimes explicitly uttered, sometimes tacitly assumed (see e.g., APA Presidential Task Force on Evidence-Based Practice, 2006; the discussion started around the start of this century and is still going strong). Tax payers, or the clients, or their parents, pay for interventions and thus they are entitled to the best possible interventions. When an intervention is offered, one presupposes that it will work or they will see results at least to a certain degree, no matter how "results" or "works" is defined and even when acknowledging that humans and human work can never be perfect. Now, EBP is based on the view that the best possible way to guarantee that something works is to scientifically prove that it works. Furthermore, EBP is seen as a foundational principle for professionals to continue to learn and maintain theoretical and practical competencies throughout their careers (Babione, 2010; Kazdin, 2008). A further assumption is that it is not only desirable but also possible to prove that interventions work (or do not work). Many of those who think EBP is important take Randomized Controlled Trials (RCTs) to be the highest level of research. In RCTs participants are randomly assigned to a treatment group or to a placebo or control group; preferably these RCTs are double blind, meaning that neither the participants nor the involved professionals know which participants are in which group.

The discussion about EBP bears resemblance to a discourse conducted in the 1970s and 80s, mainly in the field of pedagogy and education. At that time, effect measurement was not often debated as a demand of policy makers and a standard of good practice, as it is now. The discourse was rather about empirical-analytical foundations of pedagogy and education in itself (Brezinka), which brought the field of educational science closer to the natural sciences, versus phenomenological (Husserl), hermeneutical (Schleiermacher), cultural-historical (Vygotsky), or dialectical-critical (Adorno, Horkheimer) foundations, in which educational science was viewed as belonging more to the humanities, sociology, and/or political science, respectively. The dividing line also placed quantitative methods on one side and qualitative methods on the other. As I will describe, recent literature shows that in our era alternatives are discussed for what is seen by some as an undesirable narrowing of research in the social sciences with too much emphasis on causal evidence.

A related debate in the 1970s and 80s involved the so-called medical model which viewed children and adults with impairments as being ill or handicapped members of a labeled group, versus emancipatory models in which the person with a physical, sensory, mental/psychiatric, or physical deficit was viewed as an individual who was simply different from other individuals (Laing, 1971). The fact that EBP began in the field of medicine, was then adopted by psychiatry, then psychology, pedagogy, and education, is seen by some as an invasion of a medical-biological model into a practice in which individual persons, each with their own problems, wishes, characteristics, and contexts, should be central. Their individuality makes them unsuitable for empirical-analytical research (see e.g., Holmes, Gastaldo, & Perron, 2007; Thomas, Bracken, & Timimi, 2012).

Alternatives for the empirical-analytical approach of education and other fields in the social sciences, and for the supposed "medical model" are given for instance by Standal (2008), who suggests that professional practice in education and related fields should be viewed in terms of Aristotelian *phronesis* and is thus essentially indeterminate. Smith (2009) advocates a romantic approach to education, it being an essentially protean reality; he thereby reacts to what he sees as the hegemony of a typically modernist belief, i.e., that research should be scientific. Thomas et al. (2012) refer to Merleau-Ponty while arguing against what they view as technical approaches to the individual experiences of psychiatric patients, and propose a return to primary experience. They think that non-specific factors (i.e., factors that are not specific to the problem or illness being studied or treated, such as the client's values, context, relationships, etc.) are more important than specific aspects and that they are anomalies which could cause a Kuhnian paradigm change by overthrowing EBP. Gupta (2012) stated that Thomas et al.'s arguments mainly combat the medical-biological model in psychiatry and not EBP itself. However, she says that a medical/disease model of psychiatry best fits the research methods mostly used in EBP and thinks this is an undesirable narrowing of the field. All of these authors are looking for an alternative to the use of an (exclusively) empirical-analytical approach in EBP.

Several authors stress that there is a fundamental division between scientific research and its assumptions on the one hand, and professional practice on the other hand. Kazdin (2008), although generally arguing that EBP can bridge research and practice, points to the difference between professional practice and the carefully constructed scientific experiments for which patients with less severe problems and no comorbidity are selected. In actual practice, clients have problems of varying severity, receive more or rather less support from their environment, have a more or less well-educated professional at their side, etc. Furthermore, Kazdin argues that goals of treatment often differ in practice (e.g., helping the patient to function better in daily life) and in research (removing symptoms as evidence that the intervention "works"), and that even positive outcomes of research are often difficult to translate to daily functioning of patients (e.g., scoring lower on an anxiety scale after rather than before treatment does not necessarily mean that the patient's quality of life has improved). Zachar (2012) names more or less the same problems with RCTs on psychiatric patients yet he concludes that RCTs are still better than other methods in measuring the effectiveness of treatments.

Other arguments concern the relation between science and policy. Bridges and Watts (2009) discuss the broad array of questions and considerations involved in educational policy (e.g., considerations having to do with money, practical implementation, what is

attainable given the political status quo, etc.; see also Elliott, 2001; Hargreaves, 1999). They conclude that scientific evidence can at best be informative for policy and practice, not decisive. All issues in educational policy and practice are normatively "infected" and therefore the norms and values underlying policies should be made explicit and assessed on their coherence, their justifiability, and the degree to which the proposed policy is consistent with them. This plea relates to the question of what the relationship is or should be between means and ends.

Usually two positions are contrasted in this respect: first, seeing this relationship as contingent and the choice of means being determined by empirical evidence of their effectivity, and, second, seeing the relationship as internal and the choice of means being determined by ethical/moral considerations about good education (see e.g., Biesta, 2010; also discussed below; Griffiths & Macleod, 2009). However, the internal relationship between means and ends can be acknowledged and the effectivity of means can be taken into account, and these can be weighed against each other. For instance, if a financial reward is shown to prevent 12-year-olds from "binge-drinking" alcohol, but as an educator I reject financial rewards for desired behavior because I do not wish to promote a child's dependence on external rewards, I can still weigh the pros and cons of both and decide what to do. Perhaps I give the financial reward but teach the child when she is older that, in the end, external rewards are not the best motivation for good behavior. Or I deem independence of external rewards so important that I try to stimulate intrinsic motivation in the child not to drink, even though such a strategy has never been researched. Nevertheless, views on the relation between means and ends are an important issue in the EBP discussion.

Another vital question, though not always talked about explicitly, is what counts as evidence. Biesta (2010) explains that "evidence" is not the same as "truth" or "knowledge" but that it can play a part in justifying true beliefs—justified true belief being one definition of "knowledge." Oancea and Pring (2009) point to the fact that "evidence" is not the same as "proof." What counts as evidence depends on the type of research undertaken, and this in turn depends on the kind of research question that is asked. According to the authors, observable data could be "evidence" but also arguments (in philosophical reasoning) or previous judgments (in legal research).

Which research designs can produce evidence (whatever that means) is the next question, often with references to qualitative and quantitative research. Defining both is tricky. Contrasts like understanding versus explaining, subjective versus objective, non-statistical versus statistical, are too crude, and it is quite possible to combine qualitative and quantitative research (see e.g., Smeyers, 2009). Carefully worded, one might say that quantitative designs use hard observable data which can be statistically verified whereas qualitative designs focus more on understanding what they research in other ways. Oakley (2002, 2004) regrets the gap between qualitative and quantitative designs. She promotes a balanced view, doing studies in which both quantitative and qualitative designs are used, the results of which should be phrased in comprehensible language so that they are accessible for practitioners. Hanley, Chambers, and Haslam (2016) describe the history of the debate about RCTs. I refer to them for a more extensive discussion, more particularly regarding the heat of the debate. Meanwhile there are established terms denoting the difference between forms of shown effectivity: *efficacy* (i.e., an intervention

is shown to work under ideal circumstances) versus *effectiveness* (i.e., an intervention is shown to work in professional practice; Hutschemaekers, 2009; Porta, 2008).

A more fundamental philosophical approach to EBP, in which many of the objections reviewed above are implied, is given by Biesta (2010), who extensively analyzes and discards three assumptions of EBP. He applies them to education but they are applicable to all of the social sciences. He asserts, first, that the representational view of knowledge (i.e., the view that true knowledge represents how things actually are in the world) on which the arguments in favor of EBP are based is unfruitful in the context of education, if only because the experimental interventions themselves that deliver representational knowledge alter the world that they are intended to help represent. For instance, as compared to some decades ago, many people have become knowledgeable about learning disorders, using terms like "autistic" and "ADHD," whether relevant or not. This is a consequence of the popularization of the research performed on these subjects. Thus the reality that is being researched and intervened with has changed. Teachers have to hear parents' own views on the best way to treat their children's learning problems, and participants taking part in research already have (correct or incorrect) knowledge about these problems. Biesta adopts Dewey's more dynamic view that experiments and interventions deliver knowledge about relationships between phenomena instead of about "how things are." He proposes a transactional epistemology, in which our knowledge about the world and our interventions in the world constantly change that world.

Biesta's (2010) second point regards the presupposition of causal links between intervention and result. Causality assumes closed, deterministic systems whereas education for the most part is an open, non-deterministic system. Empirical research at best can prove what did work, not what will work, since variables are always changing. The relation between teaching children and their resulting learning is not the same as that between a heating iron and the resulting expansion of that iron. Thus, there are complex rather than causal links between interventions and results, says Biesta. His third point concerns the application of scientific knowledge to educational practice. Scientific studies usually take place in carefully pre-structured environments in which the complexity is reduced, for instance by allowing only participants with symptom X but not symptom Y to take part in the study, or by only using specifically trained teachers. This helps to ensure that if the study shows something "works" (or doesn't work), the intervention is the cause, and not some other variable. Since in the real world there are clients with both symptoms X and Y, and there are also less-well-trained teachers, scientific knowledge as it is can only be applied to a world that has been adjusted accordingly—in other words: to a world that is a large laboratory. Biesta implores us to consider value questions first in education. Some interventions, although shown to be effective, he says, we wouldn't use because of undesirable side effects, for instance having to take a child away from his or her parents. Only after one has decided on values, should one decide how best to achieve the goals that support these values, thereby taking into account the three problems he sketched.

Although Biesta (2010) does not say that the problems he discusses should prevent us altogether from investigating whether our interventions work, he also doesn't provide ideas about how we could do that given these problems. I think his plea for putting values first can be made regardless of whether we can or should give evidence for the effectiveness of our interventions by empirical research in an epistemologically, ontologically,

and practically sound way. Therefore, I think that the subtitle of his article: "From Evidence-Based Education to Value-Based Education," presents a false dichotomy, although I highly appreciate the plea not to forget values in striving for "evidence."

Based partly on the assumptions that Biesta rejects, Veerman and Van Yperen (2007) sketch an effect ladder of evidence in youth care. The ladder is normative; it leads to the most desired form of research and evidence. Examining it will bring me to the work of William Herbert Dray, who also developed a normative ladder, yet in the opposite direction.

Veerman and Van Yperen (2007) begin by explaining that RCTs are not always possible in educational practice for financial, methodological, or practical reasons and that the number of RCT-evaluated and proven effective interventions is still very low. Moreover, positive results are often flattered and studies often take place in controlled environments that do not resemble actual practice. Veerman and Van Yperen also discuss the flexibility of educational practice, in which interventions are usually adapted to fit with individual children or situations, which reduces the relevance of scientific studies based on laboratory experiments. With these objections, they join Biesta. However, advocating for more evidence-based interventions is their goal, and next to a top-down approach (i.e., test whether the few RCT-proven interventions indeed do work outside the laboratory) they sketch a bottom-up approach. For interventions that youth care practice has shown to be unsuitable for RCT research, other designs to investigate effectiveness should be used and, more importantly, should be valued.

Veerman and Van Yperen (2007) discuss four levels of evidence: descriptive, theoretical, indicative, and causal, which lead to potential, plausible, functional, and factual effectiveness of interventions, respectively. Each level has its accompanying parameters of evidence and types of research. RCTs are at the highest, causal level and, for instance, observational studies are at the lowest, descriptive level. Studies should take interventions in actual practice as a starting point and decide based on these which research design is feasible. Practitioners should participate in research and the interventions they develop should have clear goals and be carefully described. Thus, the gap between professional practice and scientific research can be bridged at least partly, according to the authors, by this ladder which goes from practice-based studies without causal assumptions to laboratory-based studies with causal assumptions. William Herbert Dray, almost 40 years earlier, based his ladder of rational explanation on an analysis of causality in the social sciences and the humanities.

Causes, reasons, and William Herbert Dray's ladder of understanding

Discoursing with other philosophers but mainly with Robin George Collingwood (both were philosophers of history), Dray explores different meanings of "cause" in his book *Laws and Explanation in History* (1970). At the end of his line of reasoning he subsumes these different meanings under the different functions that causality can have. Following Collingwood and pursuing his argumentation somewhat further, Dray distinguishes cause as a handle (functional for manipulating events or the behavior of persons) and

cause as a sufficient motive for an agent to act in a particular way (functional for understanding behavior of people we see as agents).

For the first type of causality, Dray gives the example of Florence Nightingale and her staff, who discovered that when they cleaned wards and medical instruments, disease rates dropped. When they stopped cleaning, disease rates increased again. It wasn't necessary for the nursing staff to understand the grounds for this connection between cleaning and disease rate (i.e., germs cause disease), Dray says; the fact that cleaning was the "handle" for letting disease rates drop or increase (all other relevant conditions being equal in both situations: ceteris paribus) made cleaning and disease to be causally connected, and not only correlated.

I would like to add to this type of causality forms of causality that we cannot handle. When I fall from a 20-meter high rooftop and land on the ground, I will break at least one of my bones. I can prevent myself from falling by being careful, but if I fall I will inevitably break at least one bone. If we call the falling from the rooftop X, and the breaking of the bone Y, we can say that X caused Y and that both "if X, then necessarily Y" (ceteris paribus) and "Y, because of X" hold true. The same pair of assertions goes for the Florence Nightingale example yet there is a difference. Suppose the germs in Nightingale's ward become resistant to the detergent she and her staff use to clean the ward; one day they might find the disease rate no longer drops when they clean the ward. So, whereas "Y, because of X" holds with similar certainty in both examples given above, "if X, then Y" is more certain (more probable) in the falling off the rooftop example than in the cleaning of the ward example. And the grounds for this is that our manipulation of reality can change reality such that the very cause–consequence relation we operate in changes (it is, of course, also possible that reality changes uninfluenced by our handling, e.g., completely new germs that are insensitive to this particular detergent turn up in Nightingale's ward; people could evolve to have unbreakable bones, but that is much less probable). This is what Biesta (2010) refers to when he discusses his first two points. Instead of the Florence Nightingale example, many examples of handle causality can be given in the field of education, physical or non-physical—see the example I gave about people being more knowledgeable about autism and ADHD.

The other type of causality that Dray (1970) discusses, sufficient motive for an agent to act in a certain way, is seen as the most relevant for historians by both Collingwood (2002) and Dray. Dray studies it quite extensively. If we wish to understand the acts of historical figures, we try to reconstruct the motive or motives that were both sufficient and necessary for them to act in a certain way (Dray, 1970, p. 122). This implies that we see these historical figures as agents, as persons, not as mere objects that obey cause–consequence empirical laws. We ascribe motives to them and these motives, although they function as causes, sometimes are reasons. Arguing along these lines, Dray, somewhat reluctantly, and citing others, comes to the conclusion that reasons can be causes. This assertion requires him to clarify the difference between causes and reasons. According to Dray, this depends on the perspective one takes. The manipulator, that is, someone who wants to intervene and change, will speak in terms of causes of behavior whereas someone who wishes to take the perspective of the agent will speak in terms of reasons for behavior (Dray, 1970, p. 154). Dray also links these two perspectives to fields

of science: history will, or should, take the perspective of the agent whereas the social sciences will take the perspective of the manipulator.

In the next section I assess Dray's (1970) conclusion and its usefulness for the field of social science and its practices. Here I first take a closer look at the "sufficient motive" type of causality and discuss the ladder of understanding Dray develops. Regarding causation in the sense of sufficient motive, "Y, because of X" holds, provided that the reconstruction we make of the agent's behavior and his or her motives is correct. However, "if X, then necessarily Y" is even less certain in sufficient motive causality than in handle causality. For by definition, if we see human beings as agents, we see them as beings who can choose their actions and thus as beings who, in most circumstances, could have acted otherwise. We might rightfully say that the reason person A left his wife was that she had an affair with his best friend and we might rightfully designate the affair as the sufficient reason, yet this doesn't mean that all husbands whose wives have affairs with their best friends leave their wives. Nor does it mean that when person A marries again and his new wife has an affair with his best friend and ceteris paribus, he will leave her. That is: not if we see person A as an agent who rationally weighs different options for action.

However, Dray (1970) puts this agent view of human beings into perspective when he explores the different degrees to which rational explanations for human behavior are possible, that is, explanations given in terms of rational reasons of agents instead of causes. He emphasizes that rational explanations are not the same as psychological explanations in terms of moods, dispositions, character traits, etc., in the actor. If an agent acted in a bad temper or under the pressure of circumstances, and didn't calculate consciously at all about the best means to use in the situation in which he finds himself, the psychologist will focus on his bad temper. Yet for the historian it might be possible, and will be relevant, to reconstruct the agent's behavior in terms of calculations of means that the agent *would* have made if he *would* have been able to make conscious calculations at the time, which led to the act he performed.

Dray (1970) uses the term "rational explanation" in a restricted sense, namely, as an explanation in terms of a rationale for the behavior in question that not only makes that behavior intelligible but also, in a sense, justifies this behavior as appropriate or "logical" for the desired end (the question of whether the end was appropriate or moral being another question, not to be answered by the historian). Dray hastens to emphasize this "in a sense" and develops a ladder for rational understanding of behavior. Behavior that at first sight seems to have no rationale may have a rationale after all. For instance, in cases where the agent was mistaken in what she took to be facts, but where her act would have been appropriate if what she took to be facts would indeed have been facts. Or in cases where the agent had certain principles or beliefs that we (nowadays) consider to be odd but, according to her particular principles or beliefs, the act was "logical."

Dray (1970) sketches a ladder from, at the one end, behavior of a historical person for which the rationale is perfectly intelligible even for people living nowadays; via behavior for which more and more extra data pertaining to the particular agent and his or her situation need to be considered in order to see that behavior as logical; to, at the other end, behavior that cannot be explained in terms of a rationale but must be explained in terms of causation.[1] For instance, a bad temper. It is a ladder more than a continuum, just like Veerman and Van Yperen (2007) do not give a continuum but a ladder, because in

both cases there is a normative presupposition involved. Whereas Veerman and Van Yperen see causal evidence as the highest level of evidence that starts, at the lowest level, with carefully describing an intervention, Dray sees turning to empirical laws as the lowest level to which we sometimes have to descend when rational explanation turns out to be impossible. Dray argues that there is a normative order in explaining human behavior, preferred both by historians and by most ordinary people: "We give reasons if we can, and turn to empirical laws if we must" (Dray, 1970, p. 138). However, his apparently neutral distinction between the agent view and the manipulator view, later in his book, makes Veerman and Van Yperen's ladder, which aims at interventions, compatible with his own ladder. That is, in the sense that it might be sensible for the manipulator but not for the one with the agent perspective to say, "We turn to empirical laws if we can and give reasons if we must."

Manipulator or agent in professional practice? A scale of causality and interventions

What are we, or what should we be, as practitioners? Should we be manipulators or should we take an agent perspective? The answer, of course, is that we have to be both, although "intervenor" would be a more appropriate word than "manipulator."

As practitioners, we intervene, and these interventions are often based on assumed causal connections between our acting and the consequential behavior of the client. Nevertheless, while practicing these interventions we usually wish to treat clients as agents, not as mere objects. We want to treat them as rational beings. And just as the historian tries to reconstruct the reasons for behavior of a historical figure by trying to put herself in the place of this historical figure, practitioners in the fields of psychology and education try to put themselves in the place of their clients, in order to understand the reasons for their behavior. Not (only) in order to better "manipulate" them, but in order to better relate to them as agents. If sometimes we "handle" clients, we do this because the end goal of our treatment is to help them to become agents, or become agents again, or become agents to a fuller degree. And sometimes we have to put the norm of treating clients as agents temporarily aside in order to reach the goal that they become agents. A similar reasoning to that I gave in the foregoing two sentences has been given with respect to the concepts of autonomy and paternalism by several philosophers. Autonomy can be conceived as a right (e.g., to make one's own decisions, unhindered by others) or as a capacity (being able to act/think/live autonomously). Paternalism towards children or towards adults temporarily lacking the capacity of autonomy whom we want to become autonomous persons (again) can be justified, among other things, when the paternalistic act helps the child or adult to become an autonomous being (again). Thus, a temporarily forced lack of autonomy helps to achieve future autonomy (see e.g., Mullin, 2014; Sjöstrand, Eriksson, Juth, & Helgesson, 2013). Manipulating clients and striving for them to become agents do not necessarily exclude each other.

The three forms of causality examined above might provide a scale[2] on which we can place interventions, or problems for which we seek interventions. This could help to see to what extent prediction of behavior is possible and thus to what extent empirical evidence for or against particular interventions might be possible.

physical causality	handle causality	sufficient motive causality
Y because of X	Y because of X	Y because of X
if X, then Y	?if X, then Y?	?? if X, then Y??

Figure 1. Types of causality for the field of social sciences.

Figure 1 is supposed to cover the field of social sciences, which is very broad. It comprises both physical causes and problems, such as education for children with cerebral palsy, and problems that can best be explained in terms of motives as causes, such as when clients with nosophobia wash their hands excessively. I have presented the forms of causality as a scale instead of giving them as discrete concepts, because in professional practice it is often very hard to draw sharp lines between them. The two examples just given lie at the far ends of the scale and are rather clear-cut. However, phenomena such as dyslexia and ADHD, for instance, are still under debate as being either an inborn brain defect or a label for persisting problems that can have a variety of causes. Not only are physical causes still not proven for many of these labels, it is also quite possible that a complex interaction of physical and non-physical factors causes these problems. Moreover, discussions regarding the consequences of defining these problems as having physical versus non-physical origins and the advantages and disadvantages of labeling clients' problems with medical terms such as "ADHD" are ongoing (see, e.g., Batstra & Frances, 2012; Ohan, Visser, Strain, & Allen, 2011).

The discussion about medical versus emancipatory models which went on in the foregoing century, although not the same, is related to these conversations about labeling children with diseases, and the growing call for EBP in my view cannot be detached from the growing number of medical labels for problems that formerly were seen as psychological or social problems. Furthermore, psychological research has sufficiently shown that human behavior varies from rather conscious, "free" behavior; to rule-following behavior (i.e., behavior framed by norms and conventions; the philosopher of education Richard Peters examined this in the context of motivation; Peters, 1963); to behavior that is so frozen that it is hard to loosen the "if event X, then behavior Y" connection; to behavior that cannot be explained in terms of reasons at all, as sketched by Dray (1970) at one end of his ladder of understanding. Finally, motives or reasons are not entirely "free" but can be heavily influenced by physical or mental states. All of this justifies presenting the different forms of causality as a scale instead of seeing them as discrete concepts.

Sufficient motive causality is not only an explanatory concept—as Dray treats it—but also a means of intervention. As practitioners, we can *provide* clients with reasons we assume to be necessary and sufficient for them to act in the way we think best helps them. In fact, Collingwood (2002) describes this sufficient motive causality as "Here that which is 'caused' is the free and deliberate act of a conscious and responsible agent, and 'causing' him to do it means affording him a motive for doing it" (p. 290). Thus, Collingwood describes sufficient motive causality as a way in which people can "cause" others to act in a certain way, although for the historian

this form of causality is only relevant in the framework of retrospective explanation. Although Dray describes this meaning of causation as given by Collingwood correctly (1970, p. 115), he doesn't draw the conclusion that thus also from a manipulator's perspective sufficient motive causality can be functional. This is an important point to which I shall return later. The fixed connections Dray makes with his ladder and his adage "We turn to reasons …" between lack of rational explanation plus "causes," contra rational explanation plus "reasons," and later in his book between causes plus manipulation, contra reasons plus agents, implies the (in my view) mistaken idea that whenever we can give rational explanations for behavior of agents, we cannot manipulate this behavior (at all).

How do interventions in the social sciences relate to this scale of causality? In Figure 2, I have tentatively plotted interventions on the scale of causality as given in Figure 1.

physical interventions	handle interventions	sufficient motive interventions
physical causality	handle causality	sufficient motive causality
Y because of X	Y because of X	Y because of X
if X, then Y	?if X, then Y?	?? if X, then Y??

Figure 2. Tentative plotting of interventions on the scale of causality.

For problems caused purely physically, we usually apply physical interventions, if available. The child who is deaf or hard of hearing can be given a hearing aid or a cochlear implant, the visually impaired adult can be given a white stick. At the other end of the scale, physical interventions are mostly inadequate. We can give the person with an excessive fear of flying a sedative when she has to fly, but if we indeed view the problem at hand as caused by a "sufficient motive" we might try to help her using, for instance, cognitive therapy.

Evidently, in all these cases the situation is not as simple as it might seem on the scale in Figure 2. The deaf or blind person's problems are not entirely solved by providing them with physical aids. The blind person needs to be taught how to use the white stick (handle intervention) and perhaps to be motivated to actually use the white stick (sufficient motive intervention). The hearing aid or cochlear implant does not completely restore the child's hearing and not every hearing-impaired child reacts in the same way to it due to different cognitive, motivational, and contextual factors. On the other hand, the person who is afraid to fly might have other contributing problems and could be helped temporarily by taking medicine in order to find the calm to work on her diverse problems. Moreover, as discussed above, most problems practitioners in psychology and education encounter are complex mixtures of physical causes, handle causes, and sufficient motive causes. It is precisely a defining feature of these fields (Biesta, 2010, points to this) that even in the most "hard" physical cause–consequence relations, interventions are helped or hindered by personal characteristics, motives, context characteristics, etc. The social sciences aren't hard sciences, no matter how much we try to make them such. A more fitting figure, therefore, would perhaps be Figure 3.

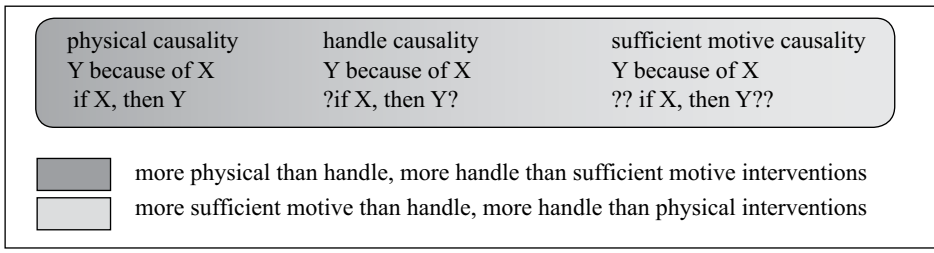

Figure 3. Scale of interventions plotted on the scale of causality.

Thus I believe the more problems and interventions are on the left side of Figure 3, the more it is possible to verify or falsify the effectiveness of interventions with randomized controlled trials or similar empirical research designs, whereas the more problems and interventions are on the right side of Figure 3, the more difficult this is. More towards the left side, causes are more clear-cut and the agency of clients plays a smaller part than it does on the right side. The relevant variables are usually more well-delineated and thus can be more easily controlled during the intervention. Teaching a visually impaired person to walk with a white stick might be an example. Even though each visually impaired person is unique, and one may be more motivated than the other to learn, the cause–consequence relations of learning to walk with a stick when not being able to see, and the relevant variables (e.g., using hearing in order to get information from the ticking of the stick to objects while walking) are relatively easy to control. Also, toilet-training children with mental impairment might be more on the left side of Figure 3. The physical cause–consequence skill (i.e., feeling the urine pressure in your bladder and then going to the potty) that should be learned is relatively simple. Methods for teaching people particular ways to handle stress resulting from an overloaded work–family combination might be somewhat more in the middle. The effectiveness of such methods can still be relatively easily researched, because the aspect of individual agency plays a relatively small role here, although there may be clients who have additional problems or reduced susceptibility to learning that interfere. On the right side are interventions that relate more to problems that have to do with psychological and rational considerations in clients. Although, as explored above, therapists can provide clients with sufficient motives to act in a certain way, the double question marks signify that the agency aspect plays an important role here. Examples might be interventions for clients with severe trauma.

If we go back once more to Dray's (1970) ladder of understanding, it might seem that I disagree with him. Dray thinks that the more we have to introduce personal characteristics in order to understand and justify a person's behavior as adequate (i.e., a person in *this* situation, with *these* assumptions about the facts, with *these* particular norms and values…) the harder it is to give an explanation for their behavior in terms of reasons and the more we need "causal explanations." According to Dray, historians "tend to push their explanations as high up 'the scale of understanding' as possible" (p. 138) yet, he says, sometimes the explanation of behavior of a historical figure needs a psychologist's analysis, where a historian must fail. And, to Dray, psychological explanations are causal

explanations, in the more empirical-law sense of "causal" that Dray employs in the first part of his book. I, on the other hand, state above that on the right side of Figure 3 we have persons as agents, with reasons, whose ability to behave otherwise than we might expect based on their own considerations makes it *more difficult* to apply empirical research to this behavior and possible interventions.

However, this contrast between myself and Dray is more apparent than true because we are dealing here with different things. As stated above, Dray treats the difference between causes and reasons somewhat inconsistently throughout his book, and his connection of reasons with historical explanation versus causes with explanation in the social sciences is too crude. Where behavior is fully intelligible in the sense that a clear, sufficient motive can be designated for it, the behavior is predictable and thus manipulable in so far as we are able to provide that motive (but note the double question marks!), whereas behavior is less predictable or at least less suitable for RCTs the more we have to deal with personal characteristics of individual persons (*this* situation, *these* norms and values…). But the main difference between myself and Dray, I think, is that he primarily comes from the perspective of how to explain Y (i.e., because of X) whereas I come more from the perspective of how to attain Y (i.e., via X), thus is the difference been a historian and a social scientist or practitioner. Nevertheless, Dray's (1970) scale is highly useful as a scale within the concept of sufficient motive causality since it indicates how far we might be able to provide persons with a sufficient motive to act in a particular way. The more intelligible the behavior in terms of rational explanations, the easier it is to provide sufficient motive. However, sufficient motive causality in itself has the lowest possibility of predicting behavior when compared to handle causality and physical causality. This brings me to my final figure, Figure 4.

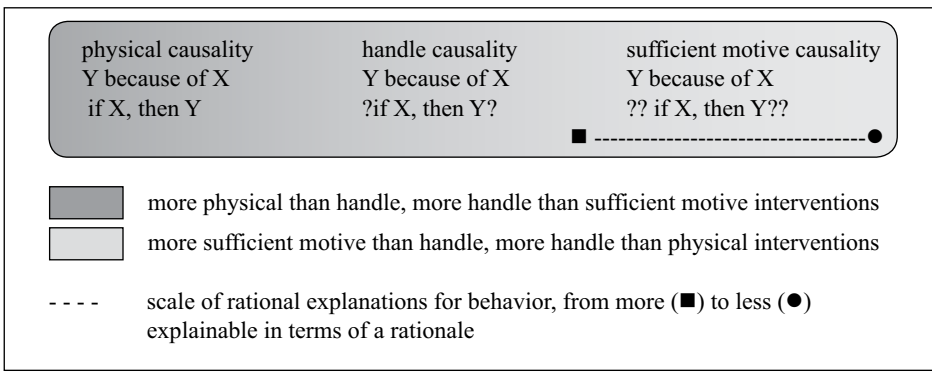

Figure 4. Scale of causality, interventions, and rational explanations plotted on the scale of causality: more realistic view.

The following rule of thumb could be used to analyze to what degree a particular intervention is suitable for "hard" research. Interventions will be more to the left on the scale (and interventions with already-proven efficacy will probably be more effective in professional practice) when the problem for which the intervention is intended is more physical (instead of more psychic) in nature; when individual variables and the context

of the client (parents, siblings, teachers, peers) play a smaller role; when the problem definition is more clear; when the problem is defined more in terms of measurable behavior; when the outcomes of the intervention are defined more in terms of measurable behavior; and when the target population for the intervention is more narrowly defined. The more to the right we go on the scale, the more other types of research are appropriate. I briefly explore these in the next section.

Quasi-experimental research designs for sufficient motive and handle causality

Interventions based on sufficient motive causality or handle causality that do not satisfy all criteria for RCTs are in principle researchable with other designs, and usually categorized as quasi-experimental research. Shadish, Cook, and Campbell (2002) state that most causes in the social sciences are in fact INUS conditions, a concept developed by Mackie (1974). An INUS condition is an insufficient, non-redundant part of a condition, the condition in itself being unnecessary yet sufficient to give a particular effect. If, for example, cognitive therapy results in decreasing the client's number of panic attacks, the therapy itself is insufficient in the sense that the client must also be cooperative, the therapist must be knowledgeable, perhaps the clients' parents play a role, etc., in order to get the effect (see the discussion of non-specific factors given above). The therapy is also non-redundant, for without it and all other conditions being equal, the effect would not occur. And assuming that another therapy might have had the same effect, the cognitive therapy also wasn't necessary.

Shadish et al. (2002) state that most studies in the social sciences investigate INUS conditions and they use this concept of "cause" to illustrate how hard it is to replicate laboratory studies and to generalize effects in the social sciences. But they also use it as a basis for their extensive discussion of many different study designs, ordered from quasi-experimental designs that least satisfy the criteria for RCTs to designs that come close to them, and finally RCTs themselves. Furthermore, they explore how quasi-experimental designs might be improved and how even the "weakest" quasi-experimental designs can have a place in the methodology of the scientist. Such designs might be used, for instance, in order to generate hypotheses about causes and effects, which helps to narrow down the variables for further studies with stricter designs. They might also be used in situations where alternative causes are highly improbable, or in which determining the exact cause of the effect is less important.

Since Shadish et al.'s (2002) pivotal book, a lot of work has been done, and is still going on, to investigate and improve quasi-experimental designs. The non-equivalent group design is probably the quasi-experimental design used most often in the social sciences. It looks much like RCT, but test group and control group are not selected randomly. They are, for instance, two classes in one school or two groups of patients in one professional practice. The error caused by the lack of randomization in this design can be repaired by applying a reliability correction (see, for instance, Riegel & Kindermann, 2016, for an example of a study with a non-equivalent group design). Kim and Steiner (2016) analyze four strong quasi-experimental designs and how they can be improved. Regression Discontinuity Designs, for instance, are suitable for situations where we wish

the clients suffering the most from their condition, or being more ill, to receive the treatment, while taking clients suffering less from their condition, or being less ill, as the control group. Although there is no randomization here of the experimental group and the control group, such a design can yield valid results. Quasi-experimental designs have also been successfully used in situations where random selection of participants is impossible and there is doubt about the generalizability of results from animal studies. For instance, in research into the effects of early risk factors (during pregnancy and perinatally) on later psychopathology (Donofrio, Class, Lahey, & Larsson, 2014). Furthermore, sophisticated types of analysis that chart the complexity of the reality psychologists and education professionals deal with have been developed, such as multivariate statistical analyses, which are especially useful in quasi-experimental designs. These techniques control for the variance in outcome measures that is attributable to differences caused by the lack of randomization of the experiment group and the control group (Anderson, 2003). And although the designs that are farthest from RCTs cannot compete with the latter with regard to validity, several quasi-experimental designs that come closer to RCTs can. They combine good inference-validity with a better ecological validity than many RCTs.

The dichotomization of research into experiments versus quasi-experiments leaves many forms of study which can differ considerably from each other in the "quasi-experimental" category. Shadish et al. (2002) distinguish between randomized experiments, quasi-experiments, natural experiments, and nonexperimental designs, whereas, for instance, Veerman and van Yperen (2007) categorize research as either causal, indicative, theoretical, or descriptive. Whatever distinction one prefers, it is important to assess the true value of each form of study, set quality criteria for it, improve it where possible, and look at what sort of intervention or what sort of situation in professional practice it is most suitable for. The scale I have developed above, and the discussion it is based on, might be of use.

Choice and justification of interventions

Now, if we accept the premise that clients are entitled to interventions that "work" as well as possible, and that it is a good professional standard for practitioners to strive for effectiveness, how then could practitioners choose and justify interventions?

Evidently, any intervention, researched or not, should conform to the usual ethical standards and should be based on clear values regarding the goal of the intervention. The possible outcomes of the intervention, including possible side-effects, and the mental and financial burden it brings for the client and her family, should be weighed against these values. A definition of the overall goal (e.g., helping clients to leave the institution and live by themselves) and the more specific goal (e.g., helping this group of clients to deal with their depression) of the intervention should be formulated before applying it. When the intervention has been researched, it should be analyzed whether the research is considered to be efficacious (it works in a laboratory situation, see above) and/or effective (it works outside the laboratory). Although these terms are used in the discussion on EBP, in research articles they are seldom used (consistently) to label outcomes. However, by looking at the type of research design used (Was it a laboratory study? What about the

ecological validity of the design?), and by placing the intervention on the scale above (perhaps using the rule of thumb I formulated), indications about efficaciousness versus effectiveness can be obtained.

If the intervention has not been researched, the degree to which it lends itself for research and, if so, for what type of research, should be analyzed. Next to the scale given above, Veerman and Van Yperen's (2007) ladder could be used here. Is the intervention clearly described? Does it have a theoretical rationale? Caution is required when the literature the intervention rests on is written by one person (or admirers of that one person), when the intervention promises to be a panacea, or when the intervention has a strong "hype"-character. The client should be informed about the degree to which the intervention has been researched and about the reasons why the intervention is proposed by the practitioner. Practitioners can help to actively lay a basis for future research of the intervention, via Routine Outcome Monitoring, that is, systematically recording treatments, interventions, methods, and their outcomes in professional practice (Hutschemaekers, 2009). This not only makes research possible, it also builds a knowledge base regarding which clients, with what problems or characteristics, the intervention does and does not work, and to what degree. For this, professional practice with its wide range of different clients seems to be better equipped than the scientific laboratory. Based on all of this, an indication can be given of how effective a particular intervention can be expected to be in the case at hand.

Concluding remarks

In this article, I have examined the potentials of scientific evidence for interventions in the social sciences. Based on the work of William Dray (1970) and some of his contemporaries, I have explored three forms of causality and developed a scale of causality, interventions, and possibilities for finding evidence in a "hard" scientific way for such interventions. I have briefly discussed less "hard" research types that can deliver proof of effectivity, and I have given some suggestions about how to choose and justify interventions, be it researched or non-researched ones. I will end with two concluding remarks.

First, of course the members of national and regional governments (depending on how the mental health care and the educational systems are organized in different countries) should be knowledgeable about what types of evidence can be obtained, in what situations, with what value, in the different fields of science. This would help get rid of the tendency to only finance interventions which are proven efficacious or effective via RCTs. Stockard and Wood (2016) investigated how many studies into the effects of educational programs and interventions, performed between 2008 and mid-2014, were included in reports of the What Works Clearinghouse (WWC). The WWC is a United States government institution that examines "what works" in education. In 2003 it set high criteria for inclusion of studies, including a focus on RCTs. Stockard and Wood (2016) report that less than two-fifths of the intervention studies performed were included in the reports of the WWC. In a second study, they looked at 131 studies of a reading program. All studies satisfied validity criteria such as those discussed in Shadish et al. (2002) and similar methodological literature. None of these studies would pass the

then strictest threshold of the WWC (a later, and different, threshold than the one of 2003), and only about one third would pass another then threshold of the WWC that was less strict. A mixed-model analysis showed that estimates of the reading program's effectiveness were unaffected by most of the WWC-threshold criteria. The authors conclude that these thresholds result in an unnecessary and detrimental narrowing of the selection of studies that are used as a basis for reviewing interventions and for consecutive policy decisions.

Second, the conception of causality as sufficient motive causality not only is a philosophy-of-science concept, useful to distinguish different forms of causality in different scientific domains. It also points to something professionals working in psychiatry, psychology, and (special) education have to deal with all the time. Psychotherapists try to motivate their clients to actively cooperate with their treatment, school administrators try to motivate teachers to implement new methods, and teachers try to motivate their students to work with these methods. And these professionals know that motivation is an important and often decisive non-specific factor, contributing highly to the effectivity of the intervention. However, the concept is much broader than that. A "sufficient motive" can be, amongst others: learning to see the irrationality of certain thought patterns (for instance by clients with anorexia); experiencing that the intervention helps to deal with severe grief; experiencing that fighting an addiction helps, and gives one a strong self-concept; learning how to: fight an addiction. "Motive" here refers more to "motion." Psychiatrists, psychologists, and (special) education providers help to set their clients or students in motion. This worthwhile activity should be backed by evidence of its effectivity where possible, yet it should be acknowledged that this is not always possible, and that different forms of "evidence" exist.

Acknowledgements

I would like to thank the editor of *Theory & Psychology* and the reviewer of the first version of this article for their very helpful comments.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Before, at the end of the book, coming to the conclusion that reasons can be causes and that it depends on one's perspective whether one speaks of reasons or of causes, Dray rather sharply distinguishes reasons from causes, and rational explanations from causal explanations.
2. Although the word "scale" implies a range from low to high, I will use this term hereafter in a non-normative sense.

References

Anderson, T. W. (2003). *An introduction to multivariate statistics*. Malden, MA: Wiley.

APA Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist*, *61*(4), 271–285.

Babione, J. M. (2010). Evidence-based practice in psychology: An ethical framework for graduate education, clinical training, and maintaining professional competence. *Ethics & Behavior*, *20*(6), 443–453.

Batstra, L., & Frances, A. (2012). Diagnostic inflation. Causes and suggested cure. *The Journal of Nervous and Mental Disease*, *200*(6), 474–479.

Biesta, G. J. (2010). Why "what works" still won't work: From evidence-based education to value-based education. *Studies in Philosophy and Education*, *29*(5), 491–503.

Bridges, D., & Watts, M. (2009). Educational research and policy: Epistemological considerations. In D. Bridges, P. Smeyers, & R. Smith (Eds.), *Evidence-based education policy: What evidence? What basis? Whose policy?* (pp. 36–57). Malden, MA: Wiley-Blackwell.

Collingwood, R. G. (2002). *An essay on metaphysics* (Rev. ed., R. Martin, Ed.). Oxford, UK: Clarendon Press.

Donofrio, B., Class, Q., Lahey, B., & Larsson, H. (2014). Testing the developmental origins of health and disease hypothesis for psychopathology using family-based, quasi-experimental designs. *Child Development Perspectives*, *8*(3), 151–157.

Dray, W. H. (1970). *Laws and explanation in history*. Oxford, UK: Oxford University Press.

Elliott, J. (2001). Making evidence-based practice educational. *British Educational Research Journal*, *27*(5), 555–574.

Griffiths, M., & MacLeod, G. (2009). Personal narratives and policy: Never the twain? In D. Bridges, P. Smeyers, & R. Smith (Eds.), *Evidence-based education policy: What evidence? What basis? Whose policy?* (pp. 115–137). Malden, MA: Wiley-Blackwell.

Gupta, M. (2012). Psychiatry and evidence-based psychiatry: A distinction with a difference. *Philosophy, Psychiatry, and Psychology*, *19*(4), 309–312.

Hanley, P., Chambers, B., & Haslam, J. (2016). Reassessing RCTs as the "gold standard": Synergy not separatism in evaluation designs. *International Journal of Research & Method in Education*, *39*(3), 287–298.

Hargreaves, D. H. (1999). Revitalising educational research: Lessons from the past and proposals for the future. *Cambridge Journal of Education*, *29*(2), 239–249.

Holmes, D., Gastaldo, D., & Perron, A. (2007). Paranoid investments in nursing: A schizoanalysis of the evidence-based discourse. *Nursing Philosophy*, *8*(2), 85–91.

Hutschemaekers, G. (2009, December). *Wetten in de weg en praktische bezwaren* [Laws in the way and practical drawbacks]. Inaugural lecture, Radboud University, Nijmegen, The Netherlands. Retrieved from https://www.lvvp.info/redactie/ckeditor422/plugins/doksoft_uploader/userfiles/files/Wettenindewegenpraktischebezwaren.pdf

Kazdin, A. E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist*, *63*(3), 146–159.

Kim, Y., & Steiner, P. (2016). Quasi-experimental designs for causal inference. *Educational Psychologist*, *51*(3–4), 395–405.

Laing, R. D. (1971). *The politics of the family and other essays*. New York, NY: Pantheon Books.

Mackie, J. L. (1974). *The cement of the universe: A study of causation*. Oxford, UK: Oxford University Press.

Mullin, A. (2014). Children, paternalism and the development of autonomy. *Ethical Theory and Moral Practice*, *17*(3), 413–426. doi:10.1007/s10677-013-9453-0

Oakley, A. (2002). Social science and evidence-based everything: The case of education. *Educational Review*, *54*(3), 277–286. doi:10.1080/0013191022000016329

Oakley, A. (2004). The researcher's agenda for evidence. *Evaluation & Research in Education*, *18*(1–2), 12–27.

Oancea, A., & Pring, R. (2009). The importance of being thorough: On systematic accumulations of "what works" in educational research. In D. Bridges, P. Smeyers, & R. Smith (Eds.), *Evidence-based education policy: What evidence? What basis? Whose policy?* (pp. 11–35). Malden, MA: Wiley-Blackwell.

Ohan, J. L., Visser, T. A., Strain, M. C., & Allen, L. (2011). Teachers' and education students' perceptions of and reactions to children with and without the diagnostic label "ADHD". *Journal of School Psychology*, *49*(1), 81–105.

Peters, R. S. (1963). *The concept of motivation*. London, UK: Routledge & Kegan Paul.

Porta, M. S. (2008). *A dictionary of epidemiology*. Oxford, UK: Oxford University Press.

Riegel, U., & Kindermann, K. (2016). Why leave the classroom? How field trips to the church affect cognitive learning outcomes. *Learning and Instruction*, *41*, 106–114.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Belmont, CA: Wadsworth Cengage Learning.

Sjöstrand, M., Eriksson, S., Juth, N., & Helgesson, G. (2013). Paternalism in the name of autonomy. *Journal of Medicine and Philosophy*, *38*(6), 710–724. doi:10.1093/jmp/jht049

Smeyers, P. (2009). On the epistemological basis of large-scale population studies and their educational use. In D. Bridges, P. Smeyers, & R. Smith (Eds.), *Evidence-based education policy: What evidence? What basis? Whose policy?* (pp. 58–81). Malden, MA: Wiley-Blackwell.

Smith, R. (2009). Proteus rising: Re-imagining educational research. In D. Bridges, P. Smeyers, & R. Smith (Eds.), *Evidence-based education policy: What evidence? What basis? Whose policy?* (pp. 175–190). Malden, MA: Wiley-Blackwell.

Standal, Ø. F. (2008). Celebrating the insecure practitioner. A critique of evidence-based practice in adapted physical activity. *Sport, Ethics and Philosophy*, *2*(2), 200–215. doi:10.1080/17511320802223527

Stockard, J., & Wood, T. W. (2016). The threshold and inclusive approaches to determining "best available evidence": An empirical analysis. *American Journal of Evaluation*. Advance online publication. doi:10.1177/1098214016662338

Thomas, P., Bracken, P., & Timimi, S. (2012). The limits of evidence-based medicine in psychiatry. *Philosophy, Psychiatry, and Psychology*, *19*(4), 295–308.

Veerman, J. W., & Van Yperen, T. A. (2007). Degrees of freedom and degrees of certainty: A developmental model for the establishment of evidence-based youth care. *Evaluation and Program Planning*, *30*(2), 212–221. doi:10.1016/j.evalprogplan.2007.01.011

Zachar, P. (2012). Evidence-based medicine and modernism: Still better than the alternatives. *Philosophy, Psychiatry, and Psychology*, *19*(4), 313–316.

Author biography

Agnes Tellings worked as a philosopher of education at Radboud University until the closing down of the Philosophy of Education Department, some 10 years ago. Since then, she researches the mental lexicon at the Behavioural Science Institute. Occasionally, she revisits her old passion and writes a philosophical article. In 2009 she published "A liberal sexual ethics for adolescence? Jan Steutel's view as a starting point" in the *Journal of Moral Education*. She also lectures on evidence-based practice.