Understanding the concept of external validity

A way to bridge the gap between research and practice?

Isabella Pistone, 2014

Abstract

The increasing demand for evidence-based practice has put focus on the gap between research and practice. The overall aim of this study is to contribute to enhanced research relevance by increasing the understanding of the concept of external validity. This thesis analyses scientific articles that discusses the concept of external validity, through the lens of ANT and its concept translation of scientific facts. The study shows that there are several definitions of external validity expressed in the scientific literature. External validity is defined as both generalisability and as a question of relevance. When problems and solutions are in focus external validity is addressed in terms of research's lack of relevance. The results of this study indicate that a way to make research more useful in practice is to give the question of external validity greater attention, in terms of methodology as well as epistemology. While further epistemological investigations is necessary for deepening the understanding of the phenomena, methodological development could possibly focus on how to incorporate more practice relevant properties early in the research process, the purpose being to create more and stronger links between the attributes of internal and external validity.

Sammanfattning

Det ökade kravet på evidensbaserad praktik har satt fokus på glappet mellan forskning och praktik. Syftet med denna studie är att bidra till en ökning av forskningsstudiers praktiska relevans genom att öka förståelsen för begreppet extern validitet. Studien är en analys av begreppet extern validitets behandling i den vetenskapliga litteraturen. Aktör-nätverksteorins begrepp översättning av vetenskapliga fakta används som analytiskt ramverk. I litteraturen diskuteras extern validitet både som generaliserbarhet och som en fråga om relevans. När problem och lösningar med extern validitet diskuteras är forskningens bristande relevans för praktiken i fokus. Resultaten av denna studie indikerar att detta problem skulle kunna minskas om frågan om extern validitet gavs mer uppmärksamhet, både i epistemologisk och metodologisk mening. Medan ytterligare epistemologiska studier är nödvändiga för en fördjupad förståelse av fenomenet, skulle metodologisk utveckling kunna fokusera på hur mer praktikrelevanta faktorer kan inkluderas tidigt i forskningsprocessen, i syfte att skapa fler och starkare länkar mellan de komponenter som avgör en studies interna respektive externa validitet.

Table of contents

1 Introduction

The gap between research and practice is a multidimensional problem, discussed both as a methodological issue, and of concern to policy development and implementation, and the evidence-based practice (EBP) movement. The increasing demands for EBP within numerous disciplines highlight the need for building bridges between research and practice (Nutley 2003), a detachment problem that has not yet been solved. One way of deepening the understanding of this gap is to investigate the tension between methodological stringency and practical relevance, as it has been shown that this tension is creating gaps between research and practice (Bohlin & Sager 2011). This tension can also be understood as a tension between internal and external validity. In an attempt to understand this problem from a new angle, this study focuses on the concept of external validity, which seems central to the debate. The scientific process of translation from the world into words is discussed by Latour (2000) and this theory is used as a tool for analysis of what should follow: a translation from words back to the world. This study examines what meaning the concept external validity is given in the literature, what problems are addressed and what solutions are suggested. In the final section, this is related to the debate concerning the gap between research and practice – an old issue that has come to fore with the increasing demand for EBP.

2 Background

2.1 Evidence-based practice

The basic idea with evidence-based practice is that methods within clinical work, practical guidelines and policy decisions should be developed out of the most reliable scientific knowledge available. EBP has grown out of evidence-based medicine (EBM), which focuses on evaluating medical interventions and EBP has then been extended beyond the field of medicine and is now an umbrella term for many different disciplines (Bohlin & Sager 2011). The Cochrane collaboration is understood as the foundation of EBP, as their production of meta-analyses and systematic reviews are seen as the core of EBP (Bohlin 2011).

The preferred study design within EBP is the RCT, which is commonly referred to as "the golden standard" (second best in the EBP hierarchy are comparative observational studies). The greatness of the RCT is the controlled study environment, which makes it possible to achieve high internal validity (Howich 2011). Internal validity is achieved when a causal relationship in a study is found to be a statistically true causal inference and no other explanation for the relationship is plausible (Shadish, Cook & Campbell 2002).

The foundation of evidence-based practice is the systematic review and meta-analysis (Bohlin 2012). The working process when making a systematic review includes systematically searching for all relevant available research, critically judging the research and then synthesising the different study results into conclusions about the effect of the intervention being reviewed. Some reviews also include a grading of the strength of the evidence supporting the conclusion (Rehnqvist 2011).

2.2   Science and technology studies

Science and technology studies (STS) is an interdisciplinary field that studies and explains the processes and outcomes of science (Sismondo 2010). A framework within STS is actor-network theory (ANT), which has developed a general social theory of technoscience. A point of departure within ANT is that science is representations of the world, and that these representations have been constructed in a process of translation: the phenomenon in focus is, in the scientific process, translated from one form (e.g. a natural object) to another form (e.g. a chemical sign) by using different material manipulations (Ibid.).

STS-theories can be used for describing, in new words, how science in a particular field works. By shedding new light over the research processes an understanding can be created that is of interest to both the scientists in the field and those outside who wants to know what is happening within the scientific field (Bohlin & Sager 2011). One of the founders of ANT is Bruno Latour. He has described the process of scientific knowledge production, how the sciences translate the world into new "stories". The aspect this thesis will focus on is *the circulating reference* (Sismondo 2010).

In an attempt to describe the production of scientific knowledge Latour (2000) has developed the theory of circulating references. He argues that scientific statements have been understood as exact copies of the world, while science is actually doing something completely different, it presents a construction of reality (Ibid). In an attempt to show how science transforms the world, Latour (2000) describes the chain of translation. This is the journey a concrete thing takes when it is moved from its context and transformed into an abstract entity in numerous intermediary steps.

3 Aim

The overall aim of this study is to contribute to enhanced research relevance by increasing the understanding of the concept of external validity. A more specific aim of this study is to provide explanatory models for how external validity affects the applicability of research.

3.1 Research questions

1. What different components do the concept of external validity include?
2. What different ideas concerning the role of science for practical activities are recognisable in the discussion about external validity?
3. What aspects of the translational process are connected to the concept of external validity?

4 Method and material

The various discussions, found in the literature, concerning the concept of external validity have been thematically analysed. The definition of the concept, problems addressed and suggested solutions have been categorised into different themes. The definitions of external validity are categorised into: 1) external validity as a property that allows research to be generalised and 2) external validity as a question of relevance. Addressed problems are categorised into: 1) Lack of relevant information in scientific publications, and 2) studies conducted under too ideal circumstances. Suggested solutions are categorised as: 1) theory-building, 2) transparency and focus on information, 3) practical clinical trails, and 4) action research.

4.1 Method

The search for literature was conducted within GUNDA (the university of Gothenburg's own search service), which includes articles, databases, books, and reports from the whole range of resources that are connected to the university. Searches were also conducted in the databases Jstor, pubmed and Cinahl continuously during the study process until I felt that all the research questions could be answered. Different search strings were used, examples of search terms are: "External validity" "definition" "problems" "solutions" "validity" "construct validity" "generalisation". To make sure that no relevant material was missed, a review over the reference lists of the literature included in the study was done.

4.2 Material

Most of the literature, found as a result from this search strategy, was from the field of public health, which means that the results from this study may be limited to this particular field. When the database searches were conducted, it was mostly the databases Jstor, Pubmed and Cinahl that were used. These databases are connected to a lot of public health journals, which can be an explanation to why the literature search in this study ended up framing this particular field. The research or "science" that is referred to within this study is therefore mainly the kind of research conducted within the field of public health, yet some studies could be placed within nursing and mainstream medicine. This study is because of this focused on the concept of external validity as defined and discussed within these specific fields. The studies in focus are experimental trials that evaluate efficacy and effectiveness of interventions. The authors, that take part in the discussions, are mostly researchers active within the field of evaluation and implementation research, i.e. fields concerned with the applicability of scientific findings.

Apart from scientific articles discussions within the Swedish council on health technology assessment (SBU) and the Cochrane collaboration concerning the concept of external validity have been included in the study, even though these did not show in the database searches. The reason for including them is that these two organisations are both leading in the production of evidence-based knowledge.

I, the author of this study, am studying a masters program in evidence-based practice at the university of Gothenburg. I have before this worked as a nurse in both hospitals, at nursing homes and within home care. I have in my previous work as a nurse been in contact with EBP, i.e. I have experiences of how EBP can work in practice both when it works and when it fails and I therefore may be prejudiced in a way that can affect the study. I have therefore been continuously aware of possible prejudice during the study in order to avoid negative effects on the study results.

There is a possibility that important aspects of the concept of external validity have not been examined, as the literature of this study is limited to the disciplinary fields of public health, nursing and mainstream medicine. It is also possible that the categorisations made by me in an attempt to understand the concept could be made in different ways, which could result in different conclusions. These aspects have to be considered when reading the results and conclusions of this study.

5   Analytical framework

In this study I use Gibbons et al (1994) theory about mode 1 and 2 science, yet the main analytical framework is Latour´s (2000) theory about the chain of translation.

5.1   Mode 1 and mode 2 of knowledge production

Gibbons et al (1994) distinguish two types of knowledge productions, mode 1 and mode 2. They describe that science in mode 1 is conducted within a single scientific discipline and problems and solutions are defined by criteria that reflect the intellectual interests of the discipline. Mode 1 is characterised by traditional cognitive and social norms that determine what counts as a significant problem, what constitutes as good science and who should be allowed to practice science. The quality within mode 1 is therefore determined by the norms of the discipline (Ibid.).

Gibbons et al (1994) distinguish a new mode of knowledge production that has developed out of the old mode 1 science, a mode 2 of knowledge production. They describe that the mode 2 science emphasises the applicability of scientific knowledge and is focused on problem solving. This production mode of knowledge is generated and sustained in the context of application, in contrast to mode 1 where knowledge is

developed first and later on applied by a different group of practitioners in a new context. A basic criterion of quality within mode 2 production of knowledge is the usefulness of scientific knowledge in the context of application (Ibid.).

This is how I interpret Gibbons et al (1994) theory about mode 1 and 2 science. I am aware that this theory includes many other aspects, but I will in this study only use the part of the theory described above.

5.2   The chain of translation

Latour (2000) describes the start of the chain of translation to be when the scientist takes a reference (a study sample) from its context. This reference is meant to represent the whole population, of whom the science would like to speak about. This is the first translation of the world. The reference that until now was part of a whole context is transformed into a concrete, decontextualized piece of material. As the translational journey continues this object leaves its material being behind and gets transportable, invariant, possible to standardise and translated into a universal code by the scientists. A reference that was previously an undefined, non-distinguishable part of a reality has now become something that can travel around the world, seemingly without changing properties. However, even though science changes the world by separating, classifying and standardise it in accordance with scientific principles (in order to find ways to understand it), something within the reference that was taken in the first step is still being preserved throughout the whole chain of translation. This is what makes it possible to trace the locality where the reference was first taken.

Latour (2000) calls the steps in this translational process for intermediary steps. In the intermediary steps the sciences transform the reality into a mixture of the original reference, a scientific discipline, human knowledge, and a particular paradigm and part of the reality is during this process transformed into a type of code that is understandable to the scientific discipline. In this manoeuvre the natural world is reorganised (e.g. in the laboratory) and patterns are created that are a mixture of science and aspects of the real world. By reorganising the new form of objects (the mixture of scientific discipline and reference) in the isolated laboratory, it is possible to identify logical patterns that can answer questions about the real world. The answer can be seen as suggestions for how to re-organise the real world in the same way as

the objects were organised in the laboratory, in order for the laboratory logic (effect) to be repeated outside the laboratory. External validity thus depends on whether it is possible to reorganise the world in accordance with the logic of the laboratory pattern. The last step in this transformation – from world to word – is the translation into words that can take the form of for example a graph or a research article. What was formerly a reality of a certain ("natural") kind has now been transformed into another, new constructed reality that is in one way more abstract, but at the same time more concrete because it is understandable to the scientific community (and maybe the society) (Ibid.).

Latour (2000) argues that every intermediary step within the chain of translation creates a rupture, and – if the chain gets interrupted somewhere in the process – it stops producing truth about that particular aspect of the original product and setting. This means that the process is reductionist in relation to those aspects of reality that are excluded during the translation process. And it means that the translational process can influence the external validity if these excluded aspects show to be vital to the sight of application. The gain of the translation from reference into a universal code is, according to Latour (2000), that the universal code can be understood by a whole scientific community and it is reproducible.

Latour (2000) explains that the process of each step in the chain of translation contains both reduction and amplification. In the process of transforming the world into word the scientists reduce properties from the world where the reference was first taken, in each step locality, particularity, materiality, multiplicity and continuity are lost. At the same time there are also amplifications or gains in each of the steps. These gains are greater compatibility, standardisation, text, and relative universality. The scientists obtain the gains by adding for example already-established practical knowledge within the scientific discipline. The phenomena being studied therefore circulate in the chain of translation in each step losing properties (reduction) and

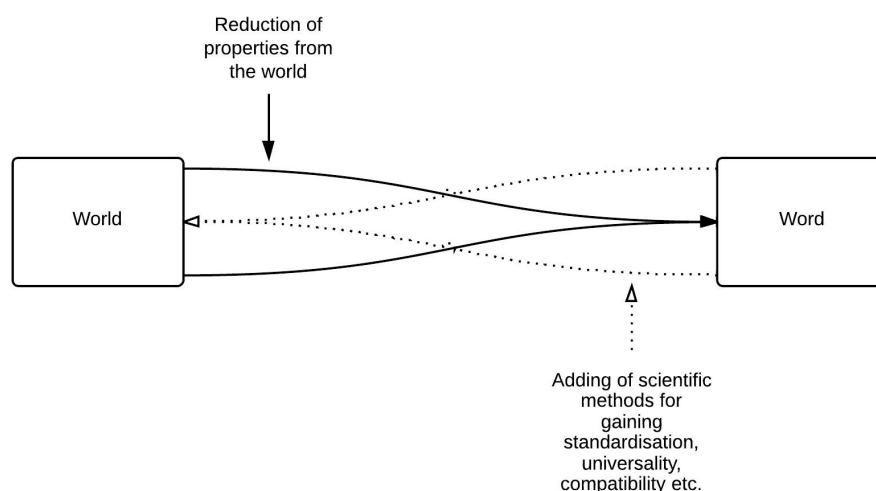gaining others (amplification) (Ibid.). This process is illustrated in figure 1.

Figure 1. The reduction and amplification as interpreted in this study.

I have now presented how I interpret Latour´s chain of translation (2000) and I will now use it to analyse the findings of this study with a focus on how the translation from world to word can be understood when the goal is to achieve external validity. I will also examine if there are different properties that are needed in order to achieve internal or external validity and if this is the case, how the transformation process or chain of translation differ.

6 Two definitions of external validity

When analysing the material for this study, I found that there are differences in the way the authors talk about external validity. In the discussions about external validity I have found what I interpret as two different ways of defining the concept: 1) external validity as a property that allows research to be generalised and 2) external validity as a question of relevance. I will now describe external validity as discussed in the literature.

6.1 External validity as generalisation

The concept of generalisation is in the current discussion about external validity (or at least the part of the discussion that this study frames) described in what I recognise as

three different categories. There are universal generalisation, generalisation to similar situations and generalisation across situations. I will explain these categories further, but first explain what Campbell and Stanley (1966) argue are the basic assumptions of generalisability and the problems with the logic of universal generalisation that underlies the discussion about the concept of external validity. I choose to start with describing what Campbell and Stanley (1966) explain are the basic assumptions behind external validity and generalisability because many of the authors in the material used in this study argue that Campbell and Stanley (1966) started the discussion about external validity by separating validity into internal- and external validity.

6.1.1 Basic assumptions of generalisation

Campbell and Stanley (1966) define external validity as a property that allows research to be generalised outside the limitations of the study. They argue that in order to fully understand the threats to external validity it is necessary to understand the basic assumptions that underlie the possibility to generalise at all. They describe that the philosopher David Hume (his explanation of universal generalisation and the problems with the logic in Hume's reasoning) is important to highlight in order to fully understand the concept of external validity (Ibid.). Campbell and Stanley (1966) do not explicate this further, so in order to understand Hume's universal generalisation I turn to Bolton (2008) and Hacking (1983). Bolton uses a theory of science perspective to explain the basic assumptions of Hume's universal generalisation and also Hacking discusses Hume as an example of the philosophy of the natural sciences.

The basic assumption of universal generalisation is that events of type A always will be followed by events of type B. Knowledge of these regular principles will enable prediction (Bolton 2008). Hume (1711-1776) was an empiricist, believed in observation, and analysed causality in terms of regular associations – or correlations – between cause and effect. What Hume described as regularities is now understood as causality (Hacking 1983). Later Mill (1806-1873) recognised that in practice what is observed cannot simply be described as type A events being followed by type B events, because in practice there are a complex of circumstances that might be affecting the A and the B. To establish the real causal link between A and B, it is

necessary to eliminate possible confounding factors C. These principles underlie our modern idea of the controlled experiment (Bolton 2008).

If we think about this in terms of Latour's (2000) concept of translation, ideally the properties that are reduced (excluded) in the translational process are the confounding factors C, i.e. the translation then reduces the properties from the world that may cause C. The properties that are added in the translation process in order to reduce confounding factors then can be thought of as the method of controlled experiment and the methodological principles that guide the modern idea of science. The result of the experiment can thus be understood as a mixture of an original reference and the adding of scientific principles.

Campbell and Stanley (1966) argue that Hume's principles of universal generalisation showed to be never fully justified logically in the beginning of the twentieth century. According to them it was recognised that it is not logically possible to generalise beyond the limitations of the study conditions (Ibid.). Despite this, Campbell and Stanley (1966) argue that a scientific field learns from the history and cumulative experience of attempts to generalise and is thereby able to justify generalisations, not logically but deducibly.

Bolton (2008) explains that in absence of the humean universal generalisation, the science tries to determine the probability of the next A being followed by B based on the sample so far observed. The probability of the next A being followed by B is then more likely to happen if the sample which the generalisation refers to is as similar as possible to the one in the study. Campbell and Stanley (1966) argue that this conclusion is based on the assumption that there is a lawfulness of nature - the closer two events are in time, space and measured value the more they tend to follow the same law. Regarding this explanation of generalisation Campbell and Stanley (1966) as well as Bolton (2008) explain the inductivist approach, where probability is a way to justify the wish to generalise scientific knowledge.

This discussion about generalisation shows that there within the scientific community seems to be questions about if it is possible to achieve external validity at all: if external validity is defined as a property that allows research to be generalised. At least it shows that it is not possible to ensure external validity, but only deducibly or

theoretically show it is probable. I will now continue to describe how external validity is described as generalisability.

6.1.2 Universal generalisations

Houlden (1980) describes that the kind of research that aims at building a theory can be called universalistic research and that it is possible to make universal generalisations when this kind of research is concerned with the testing of a grander theory. Such theory is, according to Green and Nasser (2012), often a highly generalised causal relationship that is applicable across most of the settings and populations. Universal generalisation is thus a way to generalise by connecting research to a grander theory.

6.1.3 Generalisation to similar situations

Houlden (1980) argues that besides universalistic research (that can be universally generalised because of the connection to a bigger theory) there is another kind of research that he calls particularistic research. In particularistic research, which is research conducted without the aim of testing a theory, the possibility to generalise is limited to the specific conditions that the study investigates (Ibid.). This type of particularistic research is common within the field of public health and medicine when conducting experimental trails and evaluations of interventions. When addressing external validity as generalisation to similar situations what is referred to is this kind of particularistic research.

When describing the concept of generalisation to similar situations, it is primarily Campbell and Stanley (1966) and SBU (2012) that address external validity in this way. It should be noted that Campbell in his later work also addresses external validity as generalisability across situations (Shadish, Cook & Campbell 2002).

Campbell and Stanley (1966) defines generalizability as to what extent it is possible to generalise a causal inference found in the experimental setting under a limited set of conditions, to the larger population of whom the science attempts to speak about. SBU (2012) describes generalisability by using the term *transferability* and defines it as to which extent it is possible to transfer the results found in a study to the specific conditions of the systematic review. Their understanding of external validity is that the research should be transferable to fit the research question.

The goal for SBU (2012) can be understood as research being compatible for the systematic review or meta-analysis that they intend to make. The main focus on external validity is if a study is useful to the aim of the systematic review and is thus not a concern of how generalisable the research is outside the scientific community.

The way SBU (2012) defines generalisation can be understood as what Gibbons et al (1994) distinguish as a mode 1 way of thinking of scientific knowledge, which is the traditional way of producing knowledge. When SBU (2012) defines external validity they seem to put emphasises on that research should be academically useful, i.e. according to norms for systematic reviews and the scientific society. This can be understood as a mode 1 of knowledge production that is conducted within a single scientific discipline where problems and solutions defined by criteria that reflect the intellectual interests of the discipline.

If we think of this in terms of Latour's (2000) chain of translation SBU´s (2012) definition of external validity is added in the translational process when conducting a systematic review. SBU´s (2012) understanding of the concept of external validity is integrated as part of the review method and quality criteria. There is thus a possibility that a different understanding of external validity could lead to a different review process and thus to different conclusions and interpretations of their meaning.

6.1.4   Generalisation across situations

Green and Nasser (2012) define generalisability as the concern of whether a causal relationship is applicable across various persons, settings, treatments, outcomes and contexts. They argue that the external validity of studies that aim to guide best practice is a matter of generalisability to many different populations and settings. Green and Glasgow (2006) express that efficiency trails can determine causal relationships. The question of generalisability is, in those cases, a question of whether the causal relationship is true also in similar conditions. The authors argue that if the results are going to be applicable in practice the question of generalisation across situations is important (Ibid.).

Both the Green and Nasser (2012) and the Green and Glasgow (2006) definitions of external validity mirrors the idea that scientific knowledge which aim for practical

applications should be applicable across various situations and not just to similar situations.

I have in this chapter presented how external validity is defined as a property that allows research findings to be generalised. I will now continue by presenting how external validity is defined as an entity of relevance.

6.2 External validity as relevance

Steckler and Leroy (2008) argue that relevant information about population, setting and intervention in the published articles are a prerequisite in order to decide if the research is applicable to a specific practice considering implementing it. Higgins and Green (2011) express a similar view in the Cochrane handbook, when stating that the goal of producing evidence-based knowledge is not to make recommendations about a treatment or intervention, but to provide practitioners and decision makers with enough information for them to be able to judge whether the intervention is relevant for the intended practice (Higgins & Green 2011).

Green and Nasser (2012) address external validity in terms of how applicable the research interventions are in real practice, i.e. whether it is possible to implement the intervention in terms of economic factors, organisational conditions and if the intervention is relevant for the group of people that it is intended to help within the local practice. According to them, the relevance of the research question that is asked in the first place is also included in the concept of external validity and plays a vital part of if the research is useful in practice at all (Ibid.).

This way of defining external validity is different from the definitions focusing on generalisation. When describing external validity as a property that allows research findings to be generalised the focus is on whether the research finding will be true (probable) in similar or different situations. When addressing external validity as a question of relevance, the focus is on the practical usefulness of the research findings.

The understanding of external validity as a question of relevance reflects the view that scientific knowledge should be useful in practice. This can be understood as an example of what Gibbons et al (1994) describes as mode 2 of knowledge production. The requirements of mode 2 thus seem to point the concept of external validity

towards relevance, contrasting mode 1, which directs the attention towards Humean causality issues and internal validity.

6.3 Conclusion: Definitions of external validity as expressions of the tension between different scientific goals

I have now presented the definitions of external validity found in the literature. We have seen that there are various definitions and understandings of this concept within the scientific literature. These differences can be explained with reference to the research community's varying goals of scientific knowledge production. When investigating these definitions, a tension between different scientific goals can thus be seen. When SBU (2012) describe external validity as how research findings can be translated to similar situations their main focus appears to be on scientific quality issues. Both generalisation across situations and external validity as relevance on the other hand are concerned with the applicability of research findings and it therefore reflects a view that scientific knowledge should be useful in practice.

Preciseness and simplicity, prerequisites for internal validity, too are in focus in SBU's (2012) stringent definition of external validity. Contrasting this, prerequisites for external validity such as the appreciation of and adaption to a number of variables in context of application shows in the Green and Glasgow (2006) definition of external validity. In sum, the various definitions of external validity can therefore be said to harbour the tension between methodological stringency and practical relevance that is central to EBP.

7 Scientific properties reducing external validity

When lack of external validity is discussed as a problem, the difficulties of generalising across many different variables as well as science's lack of relevance are highlighted. Lack of external validity appears to be related to applicability goals as it is mainly discussed by researchers within evaluation and implementation research, who identify this as a problem of implementation of EBP guidelines and policies. I am now going to describe these discussions, focusing on information given in scientific publications and the unnatural study setting.

7.1   Lack of relevant information in scientific publications

Glasgow, Green and Ammerman (2007) present lack of information about study populations, settings and interventions to source of external validity problems, as they identify this information as important for deciding whether the intervention is appropriate for a specific practice. Steckler and Leroy (2008) argue that the form and content of published research lack relevant information decision makers need to judge if the research can be generalised to the intended situation. Another argument is that the research articles or reviews mainly describe the effect of the intervention in comparison to another intervention with the focus on proving the causal relationship to be true within the experiment, which makes it hard to understand if the intervention is possible to implement in practice (Green & Nasser 2012). The argument is that this lack of implementation relevant information in research articles lowers the external validity and is an important reason for the failure of practice to translate research into practice.

This can be understood with the help of Latour´s concept the chain of translation (2000). In the research process relevant properties are added and reduced through out the whole chain of translation, and in this case the information about population, setting and intervention is reduced or left out in the last step of the translation process, which is the writing of the research article. This indicates that the information needed in order to implement the research in practice is reduced in the last step of the chain, but need not to be reduced earlier in the chain of translation in order to achieve stringency and internal validity. This also implies that it is necessary to add certain properties in order to generalise and apply research in practice.

7.2   Studies conducted under too ideal circumstances

Green and Nasser (2012) describe that there is often a lack of applicability to practice, even when the right information is given in the published research article. They argue that the experimental circumstances cannot be replicated in real practice. Green and Glasgow (2006) address the same problem and describe that most RCT studies are conducted under too ideal circumstances for them to be representative to the practices where the interventions are supposed to be applied. They argue that the RCT is more like an efficacy trial, a trial conducted under ideal testing circumstances, and thus do

15

not provide enough information about the effectiveness of the intervention in a broader context (Ibid.).

Tunis, Stryer and Clancy (2003) also address the problems of efficiency trails and argue that there is a need for effectiveness studies, studies that are conducted under natural circumstances. These should include more outcomes relevant for decision-makers and practitioners including a representative heterogeneous sample, cost-effectiveness and quality of life perspective.

Green and Nasser (2012) argue that most of the published studies have a non-representative sample of the population. They often eliminate those with multiple diagnoses and risk factors that cannot be controlled for in the experiment, yet in the "real world" there is no room for being this restrictive in the inclusion criterions.

Studies are usually conducted in settings over which the academic investigators have some control. Green and Nasser (2012) and Victora, Habiche and Bryce (2004) argue that while studies used in EBP are conducted under controlled and too narrow circumstances, they also improve reality by providing more training, supervision and funding compared to standard practices.

The problems described here, as research being conducted under too ideal conditions can be understood as a critique of the properties that are added and eliminated in the chain of translation. There seems to be a number of organisational factors that are added in the translation process, as a part of the research method to ensure the internal validity of the research. In addition, properties that are reduced are for example specific populations (people with multi-diagnoses), organisational factors and variables that make the research setting heterogeneous. The motive is then to ensure internal validity, but this reduction then causes problems of external validity. This because the properties necessary to reduce in order to achieve high internal validity are at the same time properties needed to achieve high external validity.

7.3 Conclusion: Usefulness as a quality criteria for scientific knowledge

In this chapter I have presented the problems with lack of external validity as expressed in the study material. The articulated problems are focused around the

problem of research not being relevant to practice, and can be understood as based on the mode 2 view of scientific knowledge production (Gibbons et al 1994). The quality criterion for research is in these discussions expressed as usefulness in practice, which implies an underlying understanding of scientific knowledge as a product useful to society. The problems addressed in the external validity debate mirrors a conflict between the traditional way of producing knowledge (mode 1) and societal demands for scientific knowledge useful to practice. The described adding of organisational factors and reduction of variables from "real-life" seems to be a problem when the goal is to implement scientific knowledge in practice. The addressed problems acknowledge a situation where properties that ensure internal validity are added in the translation process, while properties that are important to achieve external validity are either reduced or not added in the process, the results being that the research cannot be used in practice.

8   Suggested solutions to increase external validity

In this chapter I will present what I distinguish as four themes of suggested solutions to the problems of external validity. The first is *theory-building*, which is seen as a way of making research more universally generalisable. The second is *transparency and focus on information,* which is a solution that focuses on the problems described as lack of information in research articles. *Practical clinical trials* is a new kind of research method that is suggested to increase the relevance of research for the practice and *action research* is another way of conducting research with the purpose of increasing the relevance of research. Most of the suggested solutions focus on how to translate research into practice and I will again use Latour´s (2000) chain of translation to further understand what is included in the solutions.

8.1   Theory-building and construct validity

Lucas (2003) argues that theoretical knowledge is the key to generalising across populations and settings. Garcia and Wantchekov (2010) argue that the best way to improve the external validity of findings is to connect individual experiments with a theory, even when the experiments from the beginning were not theoretically grounded. Usually theory-building refers to when research is made for testing a theory

(hypothesis). The solution suggested by Garcia and Wantchekov (2010) to increase the external validity of research findings is to connect them to a grander theory. In that way it should be possible to make universal generalisations from specific research findings, even though the studies were not theoretically grounded to start with. Green and Nasser (2012) call this kind of theory-building *construct validity* and identify it as of use to reviewers within EBP when defining purpose and eligibility criteria for studies included in a review. They argue that theory can help defining how and why different participant characteristics or contextual factors can influence effectiveness and how broad or narrow the research question is.

This suggested solution connects attempts to increase external validity to the idea that research findings can and should be generalised. The end of the chain of translation is in focus, when it is suggested that the research findings are going to be connected to a theory late: just before the final translation into words. See this translation illustrated in figure 2.
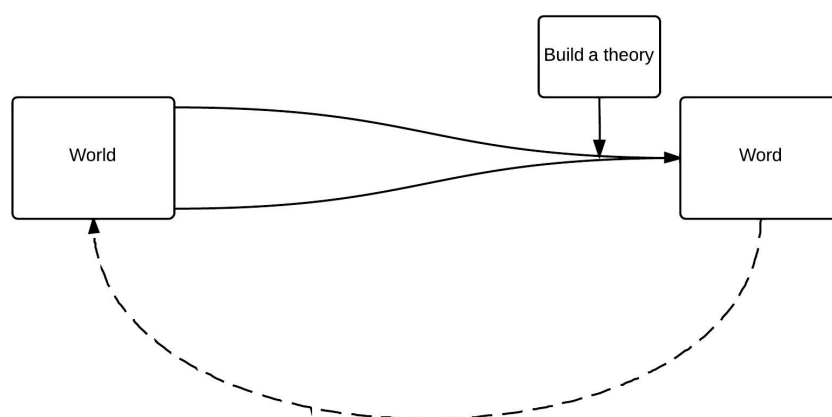

Figure 2. Theory is added in the last step before the translation into words order to make the results generalisable.

Theory-building or construct validity propose a solution to the problem of generalising specific research-findings into universal knowledge, but it avoids questions about how to practically use the theory. As the concept *theory* is not defined by Green and Nasser (2012) or Garcia and Wantchekov (2010) it is difficult to analyse what they are actually suggesting. If the basic Humean view of generalisation is recalled questions can be asked whether a norm of creating universally

generalisable knowledge is a reasonable one. If, on the other hand, generalisation is understood in a more moderate way, theory-building could be a tool for discussing a study's relevance for a specific setting.

8.2   Transparency of the research process and focus on information

Glasgow, Green and Ammerman (2007) propose a solution for increasing external validity when the problem is expressed as lack of information in published research articles. They identify four categories that can make it easier for local practices to decide if a research intervention is relevant to their practice (Ibid.).

1. Information about recruitment and selection procedures, participation rates and representativeness of intervention staff, participants and settings.
2. Consistency of implementation across settings, program components and time.
3. Impact on secondary outcomes of importance to patients, clinicians and decision makers.
4. Items from the first and third category in follow-up studies.

Glasgow, Green and Ammerman (2007) suggest that an increased reporting of external validity according to these four categories could help the translation from research to practice as it would make it easier for practitioners to decide if the study applies to their local settings, populations, staffing and resources. Green and Nasser (2012) argue that there are a number of different tools to help researchers, publishers and those who conduct systematic reviews to provide more information relevant for enhancing external validity. These vary from being a question in a quality assessment tool or reporting checklist, to validated checklists for evaluating the external validity of trails.

The solution suggested by Glasgow, Green and Ammerman (2007) and Green and Nasser (2012) is aiming for the problem of external validity when defined as generalisation to similar situations. The purpose of the solution is to make research easier to replicate by increasing the information in the research articles. This suggests that the conducted research is relevant enough to the practices. What is needed is then an increased transparency of the research process so that it is possible to imagine a replication of the research results.

Latour (2000) describes the logbook as crucial for a successful research process. He describes the logbook as what makes it possible to return to each translation in order to reconstitute its history. The suggested solution to the problem of lack of practice relevant information can therefore be seen as a call for a more detailed logbook. The logbook should describe all the intermediary steps, and Glasgow, Green and Ammerman's (2007) four categories can be seen as an identification of the intermediary steps that are necessary to express for a successful implementation of an intervention in similar settings. When they describe that information about population, intervention, settings, staff etc. should be included in the research article, the focus is again on what is included in or reduced from the research process and how this affect the final outcome of the study. The solution suggests a change in the later part of the translation process, by adding practice relevant information when writing the research article. See this illustrated in figure 3.
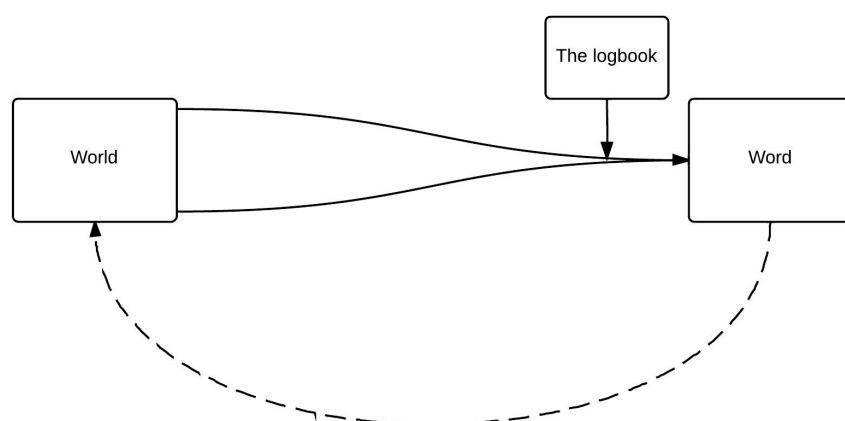

Figure 3. Information from the logbook is added in the last step before the transformation into words.

This solution can be seen as an attempt to preserve the internal validity and disciplinary scientific knowledge, at the same time as they try to increase the external validity by providing more practice relevant information in the research article. A prerequisite is that the research is practice relevant to start with and that there is no need to change anything early in the research process.

8.3   Practical clinical trials

Tunis, Stryer and Clancy (2003) suggest practical clinical trials (PCT) as a solution to lack of external validity. PCT are studies where the hypothesis and study design are formulated with specific attention to the information needed to make practice decisions. According to Tunis, Stryer and Clancy (2003) the PCT addresses practical questions about risks, benefits and costs of an intervention, as they would occur in routine clinical practice. It includes comparisons of clinically relevant alternatives, recruitment from a variety of practice settings and measurement of a broad range of relevant health outcomes (Ibid.).

Glasgow et al (2005) explain that they have developed the idea of PCT that Tunis, Stryer and Clancy (2003) describe by providing specific recommendations that should be reported in a PCT for increasing its relevance for practitioners and decision-makers. Their focus is on measurements and design choices that can enhance the external validity both regarding relevance and generalisability. They suggest a broad and representative sample from a numerous of different settings and practices. They also argue for inclusion of a heterogeneous sample regarding comorbidity (Glasgow et al 2005). They describe that the breadth and representativeness of the organisational settings (especially of the staff that deliver the intervention) is of high importance for the real world practices and therefore are important external validity properties (Ibid.).

Glasgow et al (2005) recommend that PCTs should compare clinically relevant alternatives, instead of using placebo or no treatment as a control group. This because it represents the kind of setting policy makers and clinicians is confronted with when making decisions. They also put forward that there are a need for measuring a broader set of health outcomes, economic parameters and implementation outcomes of importance to decision makers before deciding about implementation (Ibid.). As PCT is described as a study that should follow the efficiency trials or RCTs it too suggests adding practice relevant variables in the later part of the translation process that Latour (2000) describes.

Guala (2003) proposes the solution to the problem of generalisability to be to export the laboratory in different steps. She proposes that different laboratory results should

be tested in relation to different external variables. In this way the highly local laboratory knowledge can, step-by-step, be applied in reality. What Guala (2003) describes can be understood as a way of broadening the translational process by adding more real world properties to the intermediary steps. This could then result in making the final outcome more similar to context of application, while simplicity and stringency (the internal validity) are maintained. The translation process that PCTs suggest is illustrated in figure 4.
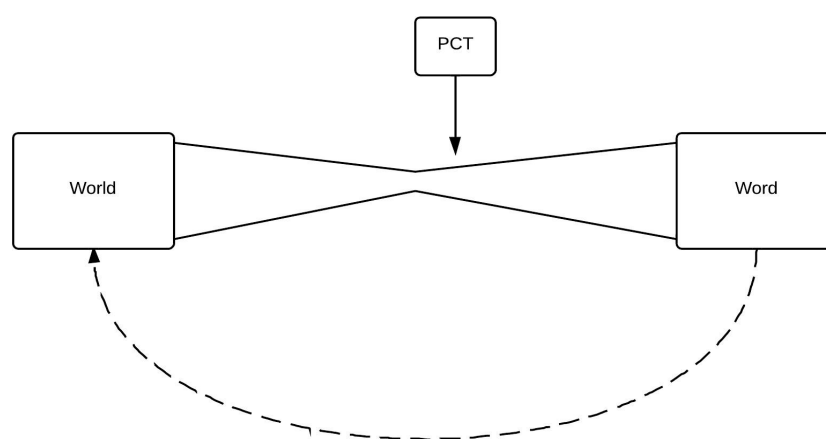

Figure 4. Practice relevant properties are added by making a PCT, after the focus on internal validity in the beginning of the research.

The basic idea is that the PCT adds practice relevant properties throughout the research process in order to "test" the links between laboratory findings, interventions and real life circumstances. The solution seems complicated and expensive and one can ask if there is an easier way to achieve the same presumed high external validity. Instead of first reducing important properties in the RCT translation process it could be investigated whether it is possible to reduce ruptures (that stop the translation and creates gaps, i.e. low external validity) earlier in the translational process. It would then not be necessary to add properties relevant to achieve external validity later in the research process, as these properties was never reduced in the first place.

While the PCT aims at solving the problems of if and how the research is going to work in practice and how it is possible to implement it, it does not solve the problem expressed as the research question and method may not be relevant to start with.

8.4 Action research

Glasgow et al (2005) describe action research as a way of conducting research that is externally valid. Action research is a form of research that starts from a question identified in practice. Glasgow et al (2005) describe that instead of top-down research, research that needs to be translated from research to practice, action research starts from the bottom-up perspective. They also argue that this kind of research is a way to bring practitioners and researchers closer together in planning, conduct and interpretation of research so that research becomes relevant to practice (Ibid.).

Brown et al (2003) further describe what action research is and explains that there are different definitions of action research. It can be defined as evaluating interventions, for raising conscience among the oppressed and as action learning. One kind of action research is called practice-research engagement (PRE). This research aims at combining the insights of practice with the analytical tools of research and is, according to Brown et al (2003) a broader definition of action research. PRE combines the definitions mentioned above with attempts to democratise knowledge by engaging in social transformation (enhancing the ability of marginalised groups to gain access to knowledge and to respect and include their knowledge). They further describe that PRE enables early identification of emerging societal problems, access to sensitive information, creation of new concepts and hypothesis, and building credibility with new populations (Ibid.).

Action research could be understood as a solution to the lack of relevance aspect of external validity. Within this research tradition methodology and research questions are formulated together with the problem setting, which are assumed to make the research more relevant. It does not, however, address the problem of generalisability, as action research is focused on creating knowledge that is highly local (A way to make it more generalisable could then be to add the suggested theory-building).

This solution is different from the other solutions as it can be understood as a circular movement between world and word, compared to the more linear movement from world to word that has been described above. The circular movement includes both theory and practice from the formulation of a research question, through planning, execution and interpretation of results. The final research results are therefore not in

needs of translation from word back to world because this translation is already a part of the process. This kind of translation process is illustrated in figure 4.
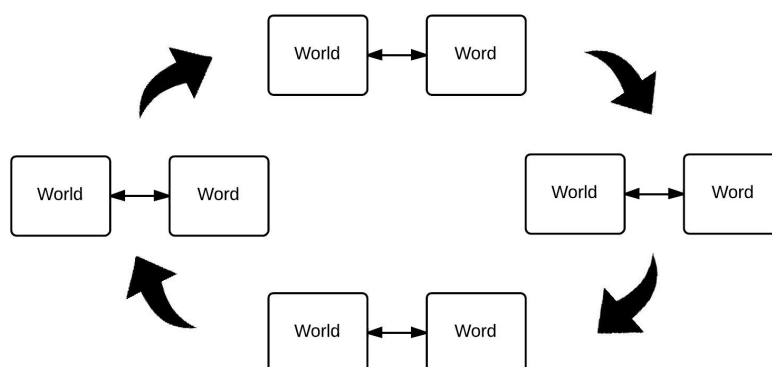

Figure 4. The translation process within action research

8.5   Conclusion: changing the chain of translation

I have in this chapter described solutions to the problem of external validity that the literature suggests and discussed them using Latour´s concept of the chain of translation (2000). The solutions approach different aspects of the problems of external validity, but they all seem to imply that there are needs for adding more practice relevant variables somewhere in the translation process or not reducing the practice relevant properties to start with. This shows that there is a need for research to be conducted differently than it has traditionally been conducted in order to be useful in practice. PCT and action research can be seen as two ways of conducting research that highlights the importance of the researchers to consider external validity also in the research process, instead of just leaving it to be a problem that belongs to the practices. Theory-building is a suggested solution to the problem of not being able to generalise research findings and can be a valuable tool for scientific disciplines that today is having problems discussing what their research is useful for.

If the above suggestions for theory-building and transparency are combined, policy and practices could be provided with both information about vital aspects of the translational process and a theory that helps linking abstract research results to clinical setting issues. When using Latour´s chain of translation (2000) as a tool to describe the suggested solutions to achieve higher external validity, it seems as

though depending on where in the translation process that the practice relevant properties are added, they solve different problems and leave others behind. When more information is added in the last step of the chain of translation (when the research article is written), the problem of lack of information to the practices is solved in the sense that this facilitates evaluation of research relevance to the specific practice. The PCT requires a bigger change earlier in the translation process and ensures external validity by not leaving this question to the practices. The solution that action research provides has a more circular translation process. More links between the world and word are maintained during the whole translation process, which should result in high external validity when defined as a question of relevance.

9 Discussion

This study shows that the concept of external validity includes a number of different definitions and that the concept is given different meanings. The concept thus seems to include a broader meaning than the Humean understanding of the concept that Campbell and Stanley (1966) described. The definitions have been categorised into two bigger themes: external validity as generalisability and external validity as a question of relevance. The multiplicity of definitions and meanings are in this study explained as differences in goals of scientific knowledge i.e. in what the purpose of knowledge is. When the goal for knowledge production is to create knowledge that develops the specific scientific discipline solving an academic question, discussions about external validity refer to stringency, internal validity and whether study findings are generalisable to similar situations. When the goal for scientific knowledge is solving a practical problem, discussions about external validity are instead focused on whether the study results are relevant to use in practice.

The basic idea of EBP is that methods in clinical work, practical guidelines and policy decisions should be developed out of the most reliable scientific knowledge available (Bohlin & Sager 2011). In the light of mode 2, (the requirement to produce scientific knowledge relevant to society) the addressed gap between research and practice that has shown visible within EBP can be interpreted as a shortcoming of the scientific practice to serve society with useful knowledge. Yet, while there is a pressure from society to make science applicable outside the universities, the research used in EBP

is still most often conducted within the mode 1 norms for knowledge production, which does not prioritise translation from research to practice.

By using Latour's (2000) theory of the chain of translation when studying the described problems with lack of external validity, it seems as the problem of relevance is connected to that not enough properties are added and reduced in the translation process. A critique of RCT studies that are highly used within EBP is that they are conducted under too ideal circumstances. This suggests that the internal validity of the studies is highly prioritised; the consequences being that practice relevant properties are reduced and other properties that help increase the internal validity are added. Looking at the addressed problems of external validity and how for example SBU define external validity this demand seems hard to satisfy using mode 1 norms for producing knowledge. It therefore seems like there is a need for new quality criteria within EBP, which today mainly focuses on internal validity criterion.

The results of this study implies that a way to bridge the gap between research and practice could be to expand the definition of external validity within SBU, and the whole EBP movement, so that it includes both generalisability and relevance. A starting point could be to look at the chain of translations: at what properties are added and reduced, when creating a systematic review for EBP. The increased focus on external validity and a deeper understanding of its role within EBP suggests it is necessary to change parts of the knowledge production process, and this study has presented a couple of suggestions for how this can be done.

While Latour´s (2000) theory of translation provides a framework for understanding the production of scientific knowledge it settles with explaining the translation from world into a new scientific understanding articulated in words. When analysing the suggested solutions to the problems of external validity it becomes clear that the chain of translation that Latour (2000) describes focuses on how to create internal validity. Developing this discussion to include the translation from word back to the world could be a way to further the understanding of the concept of external validity and its role in the scientific knowledge production.

Latour´s (2000) chain of translation has (although the focus on internal validity) shown to be a useful tool when analysing the problems and solutions with external validity, as it describes the process of knowledge production. The key for making the translation process from research to practice easier seems to be to make changes in the translation process from world to word, i.e. not to leave it to non-scientists. It has been illustrated that, depending on what kind of external validity we want, the chain of translation changes in accordance with what are added and reduced and when this happens. This study suggests that a concrete way to increase the relevance of research is for researchers to reflect on what the goal is for their scientific knowledge production, and to – thereafter – use the chain of translation as a tool for analysing what properties that are necessary to add and reduce in order for the knowledge to be useful, i.e. score high on external validity.

Within EBP this could mean that adding more practice relevant properties in the chain of translation should increase the usefulness of the research, and this study suggests that there are a number of different ways of doing this. A way to increase the generalisability of the research could be to use the study findings to build a theory to provide to the practices. A critique of EBP today is that they serve research as a readymade package, leaving it to the practices to try to squeeze some relevance out of the research (Green & Nasser 2012). In other words, EBP makes the translation from world to word, leaving it to the practices to make the translation back from word to world again. The results of this study shows that there are at least theoretically possible to make research more useful already in the translation process from world to word. Researchers would then have to plan the research process so that it becomes easier to translate the research for the practices. This study therefore propose a greater focus, on how to increase the relevance of research, already in the translational process from world to word as this could facilitate bridging the gap between research and practice.

There are other questions connected to the concept of external validity: if it is viable to produce knowledge that is not useful to the society and what useful knowledge really is. There are arguments implying that science should promote a mode 2 kind of knowledge production (Gibbons et al 1994), especially when the goal is for it to be used in clinical practice. On the other hand, it is argued that there is also a need for a

mode 1 kind of knowledge production within the universities (Gibbons et al 1994) without the pressure from society, as the usefulness of some knowledge may become obvious in a later stage. The focus of this study is on EBP where the basic assumption is that scientific knowledge should be used to ensure quality in practice. I therefore argue that research conducted within EBP should prioritise research processes that ensure knowledge that is applicable and relevant to the society.

10 Conclusion

The concept of external validity has shown to have diverse definitions depending on what the goal is for scientific knowledge production, i.e. if the goal is stringent knowledge or useful knowledge. The problems related to lack of external validity are focused on consequences of low research relevance for the practices. This study suggests that a way to increase the relevance in studies, in order to make them useful to practice, is to further investigate the concept external validity. This could deepen the understanding of the phenomena and serve as a foundation for developing research methodology accordingly. Such greater focus on external validity should focus on how more practice relevant properties can be integrated in the research process, the purpose being the creation of more and stronger links between worlds and words.

This study has limitations that could have influenced the results and conclusions. It is a very small study and there is a possibility that important aspects of the concept of external validity has not been examined and mentioned. The literature used for this study is mainly taken from the same discussions and the same authors, which have to be considered when reading the results of this study. It is also possible that the categorisation of definitions, problems and solutions made by me in an attempt to understand the concept could be made in different ways, which could result in different conclusions. Taking the limitations of this study into account, I still think the results imply that the concept of external validity is of importance for understanding the gap between research and practice. A way to proceed with the results of this study could be to empirically investigate how components of external validity is handled within a research project and to then describe the translation process with the specific focus being external validity (instead of internal validity).

References

Brown, D., Bammer, G., Batliwala, S. & Kunreuther, F. (2003) Framing practice-research engagement for democratizing knowledge. *Action research 1:*81 81-102

Bohlin, I. (2012) Formalizing Syntheses of Medical Knowledge: The Rise of Meta-Analysis and Systematic Reviews. *Perspectives on Science* 20(3), pp. 273-309

Bohlin, I. (2011) Evidensbaserat beslutsfattande i ett vetenskapsbaserat samhälle. Om evidensrörelsens ursprung, utbredning och gränser. In Sager, M & Bohlin, I. (red.) *Evidensens många ansikten.* Lund: Arkiv förlag, pp. 31-68

Bohlin, I. & Sager, M. (2011). Evidensens öde avgörs i detaljer. In Sager, M. & Bohlin, I. (red.) *Evidensen många ansikten*. Lund: Arkiv förlag, pp. 207-225

Bolton, D. (2008) The Epistemology of Randomized, Controlled Trials and Application in Psychiatry. *Philosophy, Psychiatry, & Psychology* 15:2 159-165

Campbell, DT. & Stanley, JC. (1966) Experimental and Quasi Experimental Designs. Chicago, Ill: Rand McNally

Garcia, F. & Wantchekon, L. (2010) Theory, external validity, and experimental inferences: Some conjectures. *Annals of the American academy of political and social sciences 628* 132-147

Gibbons, M., Liboges, N., Nowotny, H., Schwartzman, S., Scott, P. & Trow, M. (1994) *The new production of knowledge: The dynamics of science and research in contemporary societies*. London: SAGE

Green, L. & Glasgow, R. (2006) Evaluating the Relevance, Generalization, and Applicability of Research: Issues in External Validation and Translation Methodology. *Evaluation & the Health Professions 29:*1 129-153

Green, L. & Nasser, M. (2012) Furthering dissemination and implementation research: The need for more attention to external validity. Ss. 305-326 In: *Dissemination and Implementation Research in Health: Translating Science to Practice*

Glasgow, R., Green, L. & Ammerman, A. (2007) A focus on external validity. *Evaluation and the health professions 30:*2 115-117

Glasgow, R., Magid, D., Beck, A., Ritzwoller, D. & Estabrooks, P. (2005) Practical clinical trials for translating research to practice: Design and measurement recommendations. *Medical care 43:*6 551-557

Guala, F. (2003) Experimental localism and external validity. *Philosophy of science 70* 1195–1205

Hacking, I. (1983) *Representing and intervening: Introductory topics in the philosophy of natural science.* United kingdom: Cambridge University Press

Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org

Houlden, P (1980) A philosophy of science perspective on the validity of research conclusions: response to Anderson and Hayden. *Law and society review 15:*2 305-315

Howick, J. (2011) *The Philosophy of Evidence-Based Medicine*. Oxford: Wiley-Blackwell.

Latour, B. (2000) *Pandora's hope: Essays on the reality of science studies*. USA: Harvard College

Lucas, J. (2003) Theory-Testing, Generalization, and the Problem of External Validity. *Sociological Theory 21:*3 236-253

Nutley, S. (2003) Bridging the policy/ research divide Reflections and Lessons from the UK. *Keynote paper presented at "Facing the Future: Engaging stakeholders and citizens in developing public policy". National Institute of Governance Conference,* Canberra, Australia.

Rehnqvist, N (2011) Förord. In Sager, M. & Bohlin, I. (red.) *Evidensen många ansikten*. Lund: Arkiv förlag pp. 9-11

SBU (2012) Utvärdering av metoder i hälso- och sjukvården: en handbok. Version 2012-02-03. Stockholm: Statens Beredning för Medicinsk Utvärdering (SBU)

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

Sismondo, S. (2010) *An introduction to science and technology studies*. 11. Ed., UK: Blackwell publishing Ltd

Steckler, A. & McLeroy, KR. (2008) The importance of external validity. *American Journal of Public Health 98:*1 9-10

Tunis, S., Stryer D. & Clancy, C. (2003) Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *JAMA 290:*12 1624-1632

Victora, C., Habicht, J-P. & Bryce, J. (2004) Evidence-Based Public Health: Moving Beyond Randomized Trials. *American Journal of Public Health 94:*3 400-405