Journal for Educational Research Online
Journal für Bildungsforschung Online
Volume 12 (2020), No. 3, 3–22

Marc Bienefeld, Oliver Böhm-Kasper & Christine Demmer

Highly recommended and yet neglected: The rarity of replication studies in educational science

Abstract

Within the quantitative paradigm of social sciences, replication studies are considered of high scientific value but at the same time only a small number of actual replication studies are carried out and published. This disproportion has repeatedly been pointed out in the methodological discourse. However, while in some disciplines, e.g. psychology, there is now a growing interest in the topic, there has been no comparable development in educational science. Hence, three questions are worth considering: Can the relevance that methodology assigns to replication studies also be applied to educational science? What role do replication studies play for the current work of educational researchers? And what explanations can be explored for possible discrepancies between the designated relevance and the occurrence of replication studies? These questions determine the structure of this article: We will first discuss the methodological concept of replication studies and its relevance for educational research. In the second part of the article, the results of a qualitative explorative study will be presented. We carried out a total of 13 interviews with experts from various backgrounds in educational research, focusing on the above-mentioned questions. Beside some noteworthy similarities, the results show one main dividing line with regard to argumentation, which is the experts' personal affiliation with the quantitative or qualitative research paradigm respectively.

Keywords

Replication; Qualitative research methods; Quantitative research methods; Expert interviews; Evidence based research; Repetition

Marc Bienefeld, M.A. (corresponding author) · Prof. Dr. Oliver Böhm-Kasper · Jun.-Prof. Christine Demmer, Universität Bielefeld, Universitätsstraße 25, 33615 Bielefeld, Germany
e-mail: m.bienefeld@uni-bielefeld.de
oliver.boehm-kasper@uni-bielefeld.de
christine.demmer@uni-bielefeld.de

Extrem wichtig und doch vernachlässigt: Die geringe Präsenz von Replikationsstudien in der Erziehungswissenschaft

Zusammenfassung

*Vor allem innerhalb des quantitativen Paradigmas wird dem Replikationskonzept eine große Bedeutung zugeschrieben. Gleichzeitig werden aber nur wenige Replikationsstudien veröffentlicht. Auf dieses Missverhältnis wurde im methodischen Diskurs immer wieder hingewiesen. Während jedoch in einigen Disziplinen, wie z.B. der Psychologie, das Interesse an dem Thema wächst, gibt es in der Erziehungswissenschaft keine vergleichbare Entwicklung. Daraus ergeben sich folgende drei Forschungsfragen: Lässt sich die allgemeine theoretische Bedeutungszuschreibung auf die Erziehungswissenschaft übertragen? Welche praktische Relevanz haben Replikationsstudien im Alltag von Erziehungswissenschaftler*innen? Welche Erklärungen lassen sich für mögliche Diskrepanzen finden? Diese Fragen bilden die Grundlage des folgenden Artikels: Zunächst werden das theoretische Konzept von Replikationsstudien und deren Relevanz für die Erziehungswissenschaft diskutiert. Im zweiten Teil des Artikels werden die Ergebnisse einer qualitativ-explorativen Studie vorgestellt. Hierfür wurden 13 Interviews mit Expert*innen aus verschiedenen Funktionsbereichen der Erziehungswissenschaft bezüglich ihrer Einschätzung der zuvor vorgestellten Fragestellungen durchgeführt. Neben durchaus sichtbaren argumentativen Gemeinsamkeiten der Teilnehmer*innen erweist sich bei den Ergebnissen vor allem die persönliche Zuordnung der Expert*innen zum quantitativen bzw. qualitativen Forschungsparadigma als eine bedeutsame Trennlinie.*

Schlagworte

*Replikation; Qualitative Forschungsmethoden; Quantitative Forschungsmethoden; Expert*inneninterviews; Evidenzbasierte Forschung; Wiederholung*

1. Introduction

Quantitative-empirical research attributes great importance to the concept of replicating studies in order to provide a (more stringent) test of existing findings (replication) (see Popper, 1935; King, 1995; Schmidt, 2012). Not only in the natural sciences but also in other empirical sciences, replication emerges as “a very important method for testing empirical knowledge claims based on experimental and quantitative research” (Deutsche Forschungsgemeinschaft, 2017, p. 1). This demand for research results to be replicable methodologically is always emphasized in methodology textbooks (an example for pedagogical research is Newby, 2013). However, the concept of replication is implemented inadequately in research practice in the social sciences (see Schmidt, 2017). We can observe the radical mar-

ginalization of replication in pedagogical research in particular. A study by Makel and Plucker (2014) finds that only 0.13 % of the research results published in the 100 top-ranked journals of pedagogical research within a five-year period derived from replication studies. While there is a growing willingness to deal with the issue of replications (and associated methodological problems) in neighbouring disciplines such as psychology (see, for example Abele-Brehm et al., 2019; Camerer et al., 2018; Klein et al., 2018; Rost & Bienefeld, 2019), no similar development can currently be seen in educational science. The increasing attention that psychology is currently paying to the concept of replication largely follows the Open Science Collaboration (2015) study published in *Science Magazine* on the replicability of psychological research. This study finds that the statistically significant effects published in original studies can be reproduced in only about one third of matching replication studies. In addition, the study carries out a comparison of the mean effect sizes between original and replication studies. It finds that the mean effect size of the replication effects is only half the magnitude of the mean effect size of the original effects. The differences between the effect sizes of original and replication studies are significant here. Moreover, the authors test whether the effect sizes of the original studies fall within the 95 % confidence interval of the replication studies, which is the case in only about half of the studies analyzed (see Open Science Collaboration, 2015; controversial discussion in: Gilbert et al. 2016; Anderson et al., 2016). These findings provide the first empirical confirmation of the crisis of confidence in psychology that Pashler and Wagenmakers (2012) have postulated.

Despite the findings of Makel and Plucker (2014), a similar debate on replication has not yet been kindled in educational science. It is against this background that the present article aims to discuss the significance of the concept of replication for empirical research in educational science. We use the findings from an interview study to outline the value of the concept in the day-to-day research of educational scientists. For this purpose, we will first point out the methodological definitions and aims of replications (2.1) and relate them to educational science (2.2). We will present our interview study in which we asked research practitioners in educational science about their understanding of and experiences with replications: We will first explain the study design (3.1) and then present results from the interviews with quantitative and qualitative researchers (3.2).

2. The concept of replication and its relevance for educational science

2.1 Definitions and aims

Generally, replication refers to “the idea of reconfirming a scientific hypothesis or an experimental result by means of repetition” (Schmidt, 2012, p. 234). If we look for a universal and at the same time sufficiently nuanced definition that can serve

as the basis for actual research practice, then what emerges is more a "diffusion of concepts" (ibid., p. 235). There is a wide range of diverse and sometimes very contradictory approaches (see Cronbach, 1982; Sidman, 1960). Schmidt (2012) therefore proposes a "*functional approach*" (ibid., p. 240) that takes up the different aims of replication studies and uses them as the basis for classifying different forms of replication.

He proposes to first decide which elements should be kept constant in a replication study and which need to be altered and thereby provide a goal-oriented classification that can be transferred directly and explicitly to research practice

If, for example, the focus is on checking the sampling error, then the investigator should endeavour to keep all elements of the study constant, except for the sample. The recruitment strategy also remains the same; it is used again to obtain a new sample from the original population. If, on the other hand, the desired aim is to check lack of internal validity, then the application of the independent variable and the sample should remain constant. But as many alterations as possible should be made in the realization of the dependent variables. If certain findings should be tested for their validity for other populations, then only the population should be altered. Since all previously described aims are pursued on the basis of the same mode of investigation, such replications are referred to as "*direct replications*" (Schmidt, 2012, p. 244).

If in contrast the hypothesis is to be confirmed under completely different conditions, then the study design (e.g. sample selection, realization of dependent variables) itself has to be altered. The alterations are determined by the research question of the replication study. It is even possible to alter all elements except for the hypothesis which is then the only common basis of the two studies. This approach is referred to below as "*conceptual replication*" (see Schmidt, 2012).

Within research practice the functional approach can be seen as an ideal but hard to realize model of replication. Nevertheless, it provides a sufficiently subtle framework for replication studies to be conducted in a methodologically appropriate way, and it therefore forms the theoretical basis for the following sections.

2.2 To what extent can the concept of replication benefit educational science?

These statements show that the concept of replication is directed primarily at experimental studies and their replicability. In contrast to the natural sciences or psychology, however, actual experimental studies are scarce in educational research. This is due mainly to the structure of the field that educational science deals with: research questions in educational science usually address specific problems that occur under local conditions, and that therefore limit generalizations and the development of theories (see Berliner, 2002). He points out that research on education is generally determined by three factors that make replicability of research findings much more difficult: the power of context, the ubiquity of interac-

tions, and the dependence of research findings on processes of social and societal change (*decade by findings interactions*). Every educational activity takes place in a specific context: for example, family, kindergarten, school, or youth center. Each of these contexts has a unique structure, such as parenting behavior, municipal budgeting for pedagogical programs, leadership behavior of school heads, and how teachers teach. Due to the contextual influence of different local circumstances, study results are hard to compare. Even studies that have the same design may show different results. Here, context means above all integrating the human being in a complex and mutable network of social interactions. This becomes particularly clear if we think, for example, of teaching in a school class: the teacher has to adapt his or her actions to 20 to 30 unique combinations of student characteristics, such as intelligence, gender, eagerness and motivation to learn, socioeconomic background, and level of parental support. At the same time, student behavior is influenced by teacher characteristics, such as didactic and pedagogical competence, expectations in terms of what students can achieve, and even the well-being of the teacher at that moment. Given these complex interactions and their effect on pedagogically relevant features, the research field of educational science is much more heterogeneous than the artificial and highly controlled settings of laboratory experiments in the natural sciences and in psychology. Finally, the diverse interactions take place within their own specific social, cultural, and intellectual frameworks. Social and cultural change as well as progress in pedagogically relevant domains of knowledge (e.g. neurosciences) may cause a decade of research findings to be outdated or irrelevant already in the following decade (see Berliner, 2002).

Given this specific research field, it is worth discussing the question whether the concept of replication can be fruitful at all in educational science. We nonetheless believe that educational science should pay more attention to the concept of replication, and critically examine the robustness of its research findings. Our position is based on two lines of reasoning: first, educational science is regarded as a science of reality. Second, the idea of significance testing is central for the generalizability of empirical findings in educational science. As a science of reality, educational science seeks both to uncover regularities and to interpret an individual case embedded in specific local conditions (see Weber, 1904). However, given the complexity of the field we cannot adopt an either/or position with regard to these two aims, but instead embrace both. Educational research is characterized by such methodological diversity and by the high value placed on the discursive exchange of arguments (see Berliner, 2002). If educational researchers choose the quantitative research paradigm and apply statistical methods based on logical-deductive thinking, they inevitably face the question of whether their findings are robust and can be generalized. The above mentioned characteristics of the research field do not allow for a replication of all empirical findings. But this fact "must on no account be misappropriated as an excuse or apology for non-replicability in cases where the replicability of a scientific knowledge claim must be expected for methodological reasons" (see Deutsche Forschungsgemeinschaft, 2017, p. 3). If, as a science of reality, educational science takes seriously its aspiration to reveal also the

regularities of human actions and interactions, then we must not avoid a methodological debate on the value of replications.

Making empirical findings replicable is a demand that has always been woven into quantitative research in educational science: the idea of significance inevitably implies the idea of potential replicability, since the type I error (alpha error, i.e., the false rejection of a null hypothesis) may be minimized in its probability of occurrence, but can never be ruled out completely. Thus, a single study is not sufficient to decide whether the observation made was a clear finding or an artifact generated by error. It is at this point that the dilemma of non-experimental quantitative research becomes clear: on the one hand, the principle of traditional hypothesis testing inevitably implies the idea of the potential replicability of empirical observations, while on the other hand different contexts and complex interactions generate varying conditions for observations that cannot be replicated. One way out of this dilemma is the approach proposed by Schafer (2001): namely, to make a systematic link between replication studies and meta-analyses. Meta-analyses are not understood in this context as the statistical combination of more or less heterogeneous preparatory work. Instead, they are embedded in a research design in which replication studies are conducted in parallel in a heterogeneous research field. These replication studies serve as foundation for meta-analyses that create generalizable statements on mean effect sizes of pedagogical interventions or programs. Considering the challenges inherent in educational research (Berliner, 2002), this approach seems particularly useful and should be adopted more broadly in the discussion about a viable concept of replication.

3. Expert interviews on the value of replication studies in educational science

However, in order to enable this discussion within the discipline and beyond the purely methodological level, it seems unavoidable to deal with the following questions first: What value do educational researchers attribute to replication studies? And what reasons do these researchers give for the subordinate role of replication studies within educational research? However, as there are as yet hardly any reliable findings on this question, we decided on a qualitative exploratory approach to enable a first empirical address of this question.

3.1 Research design

Using expert interviews, we asked people who are involved in educational research in different central academic positions and who have insider knowledge with regard to our research question. Expert interviews are suitable for “reconstructing complex knowledge” (Meuser & Nagel, 2013) and for “acquiring practice-saturated

expert knowledge, the know-how of those who understand the laws and routines by which a social system reproduces, enacts and possibly modifies itself, or prevents this, but also the experiences of those who have designed and realized innovations" (ibid., pp. 457f.).

The interviewees were selected on the basis of four groups: editors of educational (research) journals, representatives of research funding, acknowledged researchers in the quantitative and the qualitative paradigm and in mixed methods, as well as researchers with experience in conducting replication studies. The latter group in particular are able to talk in an enlightening way about their experiences of the research process, which can be contrasted with the perhaps sceptical voices of other interviewees. Here, we follow the principle of a sampling that allows both minimal and maximal contrasts. In order to shed light on the conditions for and against replication studies, we take into account both individual motivations and levels of the academic system in the interviews. Of the 20 persons contacted, a total of 13 agreed to be interviewed[1] (see Table 1).

Table 1: Sample

	Researchers	Research funding	Editors	Experience in replication studies
Intended sample	8	4	4	4
Realized sample	8	-	3	2

Despite our repeated requests, we were unable to find a research funding representative willing to give an interview. As a result, this perspective and thus direct information are unfortunately not available. However, the interviewees were able to report indirectly on their experiences in obtaining third-party funding and on their activities as reviewers for research funding. In addition, our sampling technique of dividing the interviewees into four groups has not led to completely separate groups since some interviewees had double roles.

This applies above all to those who simultaneously embodied both the roles as editors and as empirical researchers. The interviewees clearly positioned themselves with regard to their own affiliation with the qualitative or quantitative research paradigm. This classification has turned out to be an important dividing line for the contrasting evaluation and therefore forms the basis for the presentation of our key findings.

All interviews were conducted by two persons, of whom at least one was a project manager of the study. The respective setting was chosen by the interviewees. The majority of interviews were conducted as face-to-face interviews in offices or

1 For data privacy reasons, we are unable to provide a more explicit description of the sample. The description only relies on characteristics classified as relevant for the questions. This is the only way to ensure that no conclusions can be made about individual persons.

meeting rooms at the interviewees' workplaces (university or research institutes). Only one interview was conducted using a digital communication tool.

Following the elaborations of Gläser and Laudel (2009, p. 111) on the concept of the expert interview, the interviews were conducted with the help of a non-standardized guideline (see appendix). The guideline was compiled in accordance with the already mentioned theoretical and empirical publications on the topic of replication studies and their dissemination in social science (see 2.1). According to the questions asked, the main focus is on personal understanding, assessment of the significance of replication studies and previous contact with replication studies. In accordance with the principle of openness, the wording and order of the guideline as well as key topics in the interviews were adapted to the respective interviewees and interview situations (Gläser & Laudel, 2009, p. 150). This was particularly useful for the interviews with qualitative researchers: Since replications are not part of the qualitative research paradigm, we expected fundamental methodological questions that could only arise from the interview situation and that among other things depended on the respective qualitative research line represented by the interviewee.

The interviews were transcribed with the help of the transcription software f4. Since the focus in the analysis is on the thematic level, the transcription is presented as normal written German (Mayring 2016, p. 91).

The analysis was carried out with the help of the qualitative content analysis technique according to Glaser and Laudel (2009). In this technique, the categories used for the analysis are first derived deductively from the theoretical preliminary considerations. When in the course of the analysis text passages containing relevant information could not be integrated into the existing framework, the already existing categories were adapted or new categories were added inductively. In contrast to some classical approaches, however, no category derived from theory was removed. This ensured that the theoretical preliminary considerations were retained during the evaluation (see ibid., p. 205). As already mentioned, the (self-) assignment of the interviewees to the qualitative or quantitative research paradigm has proved to be a relevant dividing line for the analysis. Therefore, this difference forms the basic structure for the following presentation of our findings:

We will first present the results from the interviews conducted with quantitative researchers, followed by the results from the interviews with qualitative researchers. These two parts differ in their structure. This is due to the fact that the interviews with the quantitative researchers closely followed the structure and contents of the guideline. The analysis of this group thus has a rather deductive character. The presentation of the results from the interviews with the quantitative researchers therefore closely follows the contents of the guideline (see 3.3.1). On the other hand, the interviews with the qualitative researchers already deviated from the structure of the guideline after the initial question, as we had anticipated. The qualitative researchers addressed aspects that had not been included in the theoretical guideline. In consequence, the presentation of the perspectives of the qualitative researchers on replication studies (see 3.3.2) is not based on the deductively

derived structure of the guideline, but on the inductively gained findings from the interviews.

3.2 Results from the interview study

3.2.1 Perspectives of quantitative researchers on replication studies

The interviewees whose work is mainly in the area of quantitative research methods attach as a matter of principle great importance to the theoretical concept of replication. They base their evaluation on the demand made in many methodology books for research results to be replicable. It should be noted, though, that they hold different understandings of the term *replication*.

While all interviewees mention the concept of direct replication (see 2.1), their comments differ greatly in terms of how broadly the concept can be understood. The range of possibilities that they mention extends from replications that are as similar as possible and that are expanded only by supplementary questions, to the investigation of a common hypothesis with the help of completely different research designs (Interview K, p. 1). What we notice terminologically is a recurring correspondence with the approach propounded by the American behavioral scientist Murray Sidman (1960), who distinguished between *direct* and *systematic* replications. However, none of the interviewees explicitly refers to this approach.

What they value in particular about replication studies is the fact that they can help researchers to protect the insights obtained from random influences and to generalize social mechanisms. Another potential task of replication studies in educational science is seen in the use of the secondary analysis of the original dataset to test methods of statistical analysis. In contrast to traditional secondary analyses, where already existing datasets are used for new research questions, the data are examined again in this context with regard to the same question, possibly using other methods of statistical analysis.

When it came to the question of how far the concept of replication is suitable for educational research, interviewees point out their specific research field, that is the relevance of context. While the description of this specific characteristic largely corresponds to what we have already explained in section 2.2, the interviewees nonetheless come to contrasting conclusions. Some state that replication studies are especially suited to dealing with possible changes to context. Others claim that the actual idea of replication studies is undermined here, and that for example long-term studies are a much better instrument for taking account of these conditions.

The interviewees in favor of replication studies also emphasize that the findings in educational science might not be of infinite duration but are nonetheless used to shape pedagogical practice and can therefore have far-reaching consequences. For

these interviewees, it is precisely replications that are the appropriate means to test once again whether results are still valid even after a longer period of time.

> *I'm also a bit amazed actually that it does not happen more often because I mean we all work with people. So that does not really differentiate us from medicine, and as soon as we do interventions, for example in schools, we should really say: so you are intervening in the system, you are manipulating the system in a particular way. And you have to be able to replicate; you have to show that what you are doing really has an effect. (C: 356)*

> *And in this case replication studies are just simply worth their weight in gold, if you can just give suggestions with a greater degree of certainty. How you could shape reality. (A: 544)*

3.2.1.1 Dissemination of replication studies and possible explanations

Despite their theoretical favor of replication studies, the interviewees have only had limited experience with replication studies in their actual research practice. This applies both to their own research practice as well as to their activities as editors or reviewers for research funding.

As a central reason for this, the interviewees mention the high level of pressure to innovate within the discipline, which means that they see working on areas that have not yet been researched as being more important and not least as more beneficial to their career than testing existing knowledge.

> *Because nobody wants to know: I get the same result that someone else has already found out before. So with us in the discipline, how I experience it, there's always a high level of pressure to throw new results onto the market and not to replicate. (C: 129)*

According to the interviewees, this pressure is exacerbated by discipline-related third-party funding programs, which require both innovations and action frameworks as outcomes.

Although the interviewees occasionally report that results from replications submitted for publication were not published due to their lack of innovation, most do not think that there is a general policy of rejection among reviewers and editors when it comes to publishing replication studies. Rather, they report that such articles tend not to be submitted, and stress that they would even tend to appraise such articles positively – and precisely because they themselves attribute a high value to replication studies.

> *No, not submitted. We treat them like all the others, they go quite normally into (incomprehensible: review procedure? #00:01:30#) and I would even rather, I think, be a bit more benevolent, if something like that was done. (A: 37)*

Besides the reasons already described, what the interviewees also mention as an obstacle is the research methodology of the discipline. They point out that experimental designs play hardly any role in educational research but they often play a crucial role for theoretical concepts of replication, or are even regarded as a necessary prerequisite. While some of the interviewees consider this a major obstacle for conducting such studies, others argue that replication studies should also be possible, for example, in standardized surveys since the goal is to generalize the findings made. For the interviewees, this generalization necessarily implies that results can be replicated, since otherwise a claim to generalization would inevitably be undermined. The statements made by the interviewees follow both our line of reasoning in section 2.2 on the implicit logic of replication of quantitative research in educational science, as well as the discussion of the functional approach in Schmidt (2012).

Representatives of this position also make the reservation, however, that for educational science it is above all notions of conceptual replication that are important. They put forth two questions for discussion: which conditions have to be met before a replication study can be categorized as a confirmation? Does educational science need its own concept of replication suited to its research practice? In addition, they point out that many studies in educational science have great methodological deficits that make replication almost impossible or that do not justify the effort required.

It becomes clear that the interviewees assume that the reasons they cite are interdependent and lead as a whole to the fact that a large number of those educational scientists working empirically see replications as uncreative and more of a hindrance to their career.

What is more, replications are always associated with a degree of scepticism towards existing research results. According to the interviewees, this methodologically justifiable scepticism is associated especially by junior researchers with the concern that they be regarded as "traitors" and block crucial career opportunities by testing or refuting published findings.

3.2.1.2 Opportunities to promote replication studies

If the aim is to raise the status that replications have in educational science, then according to the interviewees there are different ways of doing so.

Here, the interviewees consider it indispensable to raise awareness for methodological problems within educational science. They argue that only a stronger integration of the concept and value of replication studies in university courses, con-

ferences and publications can help to develop an understanding of which aims and benefits are associated with such studies. Moreover, a discipline-specific concept of replication should also be developed, one that properly addresses and takes into account the methodological reality of research in educational science. In order to promote this discussion, the interviewees consider it important that disciplinary associations and other academic organizations such as the Science Council participate and take a clear stance.

However, since university research today heavily relies on third-party funding, the theoretical debate can only be one component in the promotion of replication studies. Those providing research funding could help if they appraised projects by not only seeing innovation as being particularly important, but also by taking into account the value of replication studies, and if they ideally even established funding lines or an "award for replication studies" that specifically promote such projects.

> *Umm then / I think what is also needed is an impulse from those who fund research. That is those who distribute the money have to, so you can see that: as soon as a funding line is built up by politicians, for example, or by science policy, then it happens. So it also takes a bit of a push from the outside, because it's very time-consuming. (C: 297)*

Reserved sections in journals and relevant calls for papers could also help promote replication studies.

The interviewees expect that these ways of promoting replication studies could enhance one another. For example, it seems entirely plausible that a larger number of specifically funded research projects will help to ensure that more such results are submitted and published. This, in turn, has the potential to show young researchers that replication studies are a key component in verifying results empirically, and do not stand in the way of an academic career.

3.2.2 Perspectives of qualitative researchers on replication studies

In qualitative research, replications are not a criterion of quality and are hence not discussed in methodology books. Nevertheless, if we include qualitative research in our study, then we are by no means implying that there is an unfulfilled methodological claim for such studies. Rather, it is a matter of exploring where qualitative research has potential links with and differences from the described functions that replications have for the logic of quantitative research. Thus, the line of questioning pursued by our project shifts when it comes to qualitative research:

- To what extent do qualitative researchers regard the replication of studies as a positive option (or not) for qualitative research?
- (Where) do qualitative researchers see elements in the qualitative research process that bear a connection with replication?

- (To what extent) do they think the replication of studies could open up interesting perspectives for qualitative research?

Unlike the more uniform premises of quantitative research, empirical approaches in qualitative research are characterized by a considerable variety of methodological principles. The interviewees are located in different schools of qualitative methodology, so that a certain range of research approaches such as ethnographic, as well as different reconstructive and interpretative approaches are represented in the sample[2]. One interviewee had experience in conducting qualitative replications.

All interviewees assign the idea of replication in the narrower sense to the quantitative research paradigm and to approaches shaped by the natural sciences, such as medicine, physics, and also pedagogical-psychological research. What the interviewees identify as being typical of these approaches is a nomothetic research logic that aims for statistical representativeness and statements with general validity. Replications are seen here as a means to test the validity of results and research tools. This results in a clear demarcation line of qualitative-interpretative research from replications in the narrower sense. This demarcation is justified methodologically, but also – and this is particularly interesting for our research question – in relation to the educational research field (see 2.2). However, when they were presented with broader approaches of replication (see 2.1), some of the interviewees certainly saw worthwhile and potentially transferable aspects of replications for their own work:

> *Well, I would have said spontaneously: 'This is a problem for quantitative researchers and I have absolutely nothing to do with it'. But that is not true ... [it's] not as simple as that. (E: 12)*

3.2.2.1 Methodological objections to replications from the perspective of qualitative research

The interviewees consider the principle of replication, whereby existing results are tested and checked, to be diametrically opposed to the fundamental premise regarding the openness of qualitative research. Thus, the research process is based on an essentially open-ended question. This research process may well comprise initial settings, but these are settings that act as theoretical heuristics and are intended to be open to dispute during the research process. Thus, the original question may shift again in line with the circular research process – something that is not compatible with the stringent testing of given results. While replications require that initial settings be kept constant, our interviewees believe that qualitative

2 We did not interview researchers who place themselves methodologically within qualitative content analysis. Since qualitative content analysis is originally an approach of quantitative analysis, and quantifying logic is often used with the approach in qualitative projects, closer proximity to replication studies would be expected.

research uses dynamic concepts. Among other things, these concepts are constituted by the unique structure and logic of the respective research field and research object, which emerge successively. Assumptions that reduce complexity, as they are needed for variables in quantitative studies (e.g., in the form of a variable 'gender'), are therefore not set from the beginning in qualitative research. These presumptions are shifted into the research process itself in order to be able to open up complexity in the first place (e.g., "*What form does gender take here?*") (M: 26).

The interviewees also describe the value and handling of contextuality, as well as the relationship between subjectivity and standardization, in a very different way. Drawing on the theoretical supposition that research objects are historically, culturally and socially embedded, they argue that qualitative research is always concerned with concrete and complex cases. These cases are systematically considered in their contextual embedding under "*field conditions*" (D: 9), which opposes the idea of the replicability of results (*ibid.*). For the interviewees, influencing contexts should not be controlled and standardized in qualitative research, but rather firmly integrated. This is the complete opposite to experimental research, which is considered the classic area where replications are used. For in experiments, contextual factors as well as the subjectivity of researchers are kept as low or as constant as possible. In contrast, qualitative research usually incorporates natural elements of the data collection process. The interviewees conceive the data collection itself as an instance of social interaction that explicitly welcomes the unfolding of the researcher's subjective relevance. In order for this to emerge within research in the first place, this logic requires not as little but as much contextuality as possible, which is again considered contrary to replication: "*You have to have an idea of the high validity of context to want to replicate something*" (M: 29).

Unlike in quantitative approaches, the interviewees surmise that the knowledge generated is subject to "*probation*" (L: 14), but less so to rules and let alone to laws. Thus, according to the interviewees, there is in some strands of qualitative research the claim to generalization, but no claim to truth. Proven knowledge can become obsolete in a changing field and is only valid for the time being. It may therefore also make sense for qualitative research to investigate whether the knowledge generated is still valid. This may suggest a point of contact with the theoretical foundations of quantitative research and with replications. It entails, however, a new and fundamentally open, as well as theoretically grounded, research question – and less the pure testing of knowledge already gained (L: 16).

Thus, qualitative research is concerned less with making the scientific knowledge generated more reliable or with quantifying it, and more with exploring its complexities, such as when a particular research question is worked on anew: "*That's why I think that the basic figure is more difference than identity. And you would, I think, that is the idea of the research gap, do exactly what you can expect new knowledge from, and not what confirms existing knowledge*" (B: 46; M: 18).

The understanding of "*research gaps*" mentioned here is also cited by the quantitative researchers and points to a common understanding of science or a shared

scientific system in which the aim is to generate new knowledge and where there is sometimes reported a great pressure to innovate. However, the discovery of new connections and the generation of new theoretical concepts is part of the original self-understanding of many qualitative research approaches, so that this also touches on a core area of methodology.

Ethnographic researchers also emphasize that the value of methodology or of the approach underpinned by methodology varies within qualitative research approaches. Thus, the quality of an ethnographic study can only be guaranteed to a limited extent by means of a determinable methodological approach. Instead, reconstructive approaches such as the documentary method establish clearly structured evaluative steps and thereby demand a more standardized methodological approach.

3.2.2.2 The understanding of the research object and the research landscape of educational science

From the interviewees' point of view, the research objects of educational science cannot be studied separately from their respective concrete, complex and ever-changing contexts: "*We look at something different each time*" (L: 1). Therefore they are sceptical of replications aimed at testing the research tool and of mere adaptations of the debate on replications from other disciplines in which research objects appear less variable. Moreover, for the interviewees, other disciplines feature more uniform research approaches. In contrast, educational science is characterized by a distinctly differentiated research landscape that, with its heterogeneity of approaches, is not – nor has to be – universally compatible with the claim to replication. They clearly reject restricting educational research to 'evidence' because the concept taken from medicine and linked to replication does not do justice to the area of interest of educational science and to its heterogeneous research landscape.

3.2.2.3 The value of research for pedagogical practice and pedagogical professionalization

What may also be specific to the discipline is the argument that science and research do not directely provide recommendations for action because they do *not have the role of justifying practice, but rather of opening up possibilities for practice to reflect and gain a certain distance* (B: 42). Hence, research has the task of providing complex insights and concepts that enable pedagogical professionals to develop the ability to condense complexity into concrete decisions for action. Unlike in laboratory situations pedagogical practice is characterized by uncertainties and by influences that are hard to control. For that reason, there can in essence be no reliability of application that is, no guarantee that knowledge from educational science can be successfully put into practice. That also applies to rep-

licated knowledge and models. Thus, the "technology deficit" (Luhmann & Schorr, 1982) of pedagogy runs counter to the idea of a knowledge that is sufficiently verified through replication and that is therefore technologically applicable. According to the interviewees, disciplines in which replications have a high status are aimed more at adding value through practical application than at adding value through knowledge. For the interviewees, they are oriented less towards activity that prioritizes understanding and abductive reasoning, and more towards a "*what-works logic*" in practice.

3.2.2.4 Frictions in research policy and research psychology

The experts in qualitative research also say that replications in some ways run counter to the research system, which focuses primarily on innovation. A "*certain compulsion to originality*" is observed (E: 10). When it comes to a scholar's own academic reputation and the success of his or her academic career, the interviewees also report considerable internal university pressure to generate third-party funding, and they find replications not helpful for this purpose. They also suspect that the publication policy of journals is not geared towards replication studies. Another hindrance are the political ambitions in which science is entangled. Especially for educational science they observe an unreflected actionism on the part of policy-makers who draw on research results and show little desire for lengthy and thorough proofs even in quantitative studies.

In addition, the qualitative researchers suspect that there is a "*state of truce*" (B: 26) among researching colleagues in which, in a figurative sense, everyone likes to "*tend their own garden*" (B: 12; also, J: 11). For the interviewees, the repetition of studies may also reveal potential inadequacies in the initial study that contradict the research claim, but nonetheless arise in practice (J: 11). Replications could therefore be considered a vote of no confidence, and could make a person seem like lacking loyalty amongst colleagues (D: 23). There is the suspicion that such a vote of no confidence is even more automatically directed at the researcher in qualitative research than in quantitative studies. This is because the entire research process and findings in qualitative research are more closely tied to the researcher, his or her methodological decisions and interpretations, while in quantitative research many research activities are standardized components of research tools (M: 34).

3.2.2.5 Potential analogies and links to conceptual replications in qualitative research?

Adopting a broad interpretation of the idea of replication, the interviewees also see similar mechanisms of testing or review in some aspects of qualitative research, for example in theoretical sampling. In this case, the sample is first drawn according to theoretical assumptions, but then successively and contrastively on the ba-

sis of initial results. Here, the following question is central: Are there similar cases that contradict the previous findings or could make them more nuanced? The sample is complete if other cases no longer bring new results to light, but only confirm the results already found. This step defines the study's scope of application, so that an analogy to the idea of replication can be seen (M: 14, E: 3). However, the interviewees also emphasize the factor of differentiation in this research step, which is directed more at investigating the variability within the phenomenon observed than at finding regularities across different contexts (D: 12).

The closest proximity to conceptual replication is attributed to research that aims at (theoretical) generalization (D: 7-8): "*Structural generalization could actually be open to replication*" (M: 9). In addition to theoretical sampling, additional factors that validate and therefore 'replicate' can be identified here. For example, in the procedural step of creating interpretations within objective hermeneutics (see Oevermann, 2000) (B: 36), where alternative meanings are developed in the discourse and examined along the case until one interpretation establishes itself and no interpretations that are more valid can be found (M: 9ff.). Since the case structure hypotheses thus formed can only claim validity until further notice, it would certainly be possible for another research group to 'test' these structure hypotheses using the same or indeed different material. However, the interviewees state that this only happens to a certain extent at conferences, where results are presented and discussed. This can be seen as a curtailed form of the 'plausibility check' (M: 5). At the same time, probation in academic discourse is seen as an important form of critical (self-)examination: 'validity' of results is measured in the reception and discussion of the results by other academics, by "*usefulness [for subsequent research], less by evidence*" (B: 9, 22, 37). In order to allow readers a critical methodological understanding and to take into account the quality criterion of transparency, the corresponding data material has to remain accessible (E: 6) – which, for example with regard to already archived material, immediately raises questions about anonymization, data privacy and research ethics (J: 11).

4. Summary and conclusion

In summary, a large proportion of those interviewees belonging to the quantitative paradigm attribute (in line with the methodological literature) a clear value to the concept of replication when it comes to safeguarding empirical research results. In contrast, those interviewees belonging to the qualitative paradigm see no direct links to replications in the narrower sense. Only one interviewee sees replications that have the aim of testing qualitative studies as being a potentially viable, but at the same time quite pointless, approach. On the other hand, the qualitative researchers deem the repetition of research questions and studies with deliberate variations (in particular, variations of sample, of historical-temporal context, of methodological or theoretical approach) as certainly being useful. They see the

aim of replication here as being to explore the socially or historically altered field of research, or to cast a new perspective on results in the light of current methodological and theoretical questions. However, in order to be able to deal with a certain research question in a comparative manner, e.g., a study presented forty years ago, it is necessary to adapt various facets of the research design, since both the research field and the methodological discussion have changed. While some of the interviewees see the concept of replication as an opportunity here, others argue for a clear conceptual demarcation to be made so as not to obscure the different methodological premises of quantitative and qualitative research approaches. Since preconceptions of replications are rather narrow, introducing replication into qualitative research contexts would lead to misconceptions about the aims and premises of qualitative research.

Despite attributing a very different value to replications, the researchers belonging to the quantitative paradigm also agree with the statement that the object of research in educational science is much more dependent on context than is the case in many disciplines in the natural sciences. However, this does not lead to the belief that replications cannot be applied in educational research and are of no benefit. Rather, it is precisely replication studies that may be a suitable means for enabling researchers to confirm or refute results under altered conditions. Due to the contextuality and other limitations of the research practise of educational science (e.g., limited research resources, changing staff, few experimental designs), the researchers in the quantitative paradigm also see direct replications as being barely feasible. On the other hand, most quantitative experts assign a significant value to the approach of conceptual replication.

The experts put forth some shared arguments for the low prevalence of replication studies. Thus, researchers from both research paradigms report of structural reasons that are in line with the DFG's position on the replicability of research results, which states: "The weight of quantitatively-parametrising control, evaluation and gratification systems prevailing in research today has the effect of creating increased (and continually increasing) pressure to compete and to accelerate results. This is manifested in decisions (and underlying decision criteria) about career moves, financial support, location of publication, and institutional structural trends" (Deutsche Forschungsgemeinschaft, 2017, p. 4). The interviewees speak in this context primarily of an enormous pressure within the discipline to publish and be innovative, a pressure that is increased not least by a system of research funding that places a high value on generating "new" knowledge, and that links support to a presentation of results that can be used directly for pedagogical practice. According to all interviews, this affects how replication studies are perceived within the discipline, with replication studies being deemed, and especially so by young researchers, as not being particularly career-enhancing.

According to the quantitative researchers, who largely accept the concept of replication, replication studies and their reputation in educational science should be specifically promoted in the future. What is required for this purpose is a deeper and broader theoretical discussion of the fundamental concept as well as the possi-

bilities of conducting replications (for example, in university teaching, and at conferences). But, an expansion of relevant third-party funding lines is also required. Such a process can be supported for example by editors of empirical journals that promote the publication of such results, or by academic associations that promote such studies.

This is different for qualitative research. Following the assessments of the experts, the specific functions of replications and the associated methodological challenges and consequences must be discussed first. It would be interesting to look at inherent understandings of 'replications' and theoretical generalization in the sense of typology or structure hypothesis within different qualitative approaches.

References

Abele-Brehm, A. E., Gollwitzer, M., Steinberg, U., & Schönbrodt, F. (2019). Attitudes toward open science and public data sharing. A survey among members of the German psychological society. *Social Psychology 50*(4), 252–260.

Anderson, C. J., Bahnik, S., Barneztt-Cowan, M., Bosco, F. A., Chandler, J., Chartier, C. R., & Zuni, K. (2016). Response to comment "Estimating the reproducibility of psychological science". *Science*, 351, 1037-c. https://doi.org/10.11126/science.aad9173

Berliner, D. C. (2002). Comment: Educational research: The hardest science of all. *Educational Researcher, 31*(8), 18–20.

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., & Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behavior, 2*, 637–644.

Cronbach, L. J. (1982). *Designing evaluations of educational and social programs*. San Francisco: Jossey-Bass.

Deutsche Forschungsgemeinschaft. (2017). *Replizierbarkeit von Forschungsergebnissen. Eine Stellungnahme der Deutschen Forschungsgemeinschaft*. Deutsche Forschungsgemeinschaft: Bonn.

Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016). Comment on „Estimating the reproducibility of psychological science". *Science, 351*, 1037-b. https://doi.org/10.1126/science.aad724

Gläser, J., & Laudel, G. (2009). *Experteninterviews und qualitative Inhaltsanalyse, als Instrumente rekonstruierender Untersuchungen* (3., überarbeitete Auflage). Wiesbaden: VS.

Luhmann, N., & Schorr, K.E. (1982). Das Technologiedefizit der Erziehung und die Pädagogik. In N. Luhmann & K.E. Schorr (Eds.), *Zwischen Technologie und Selbstreferenz. Fragen an die Pädagogik* (pp. 11–41). Stuttgart: Suhrkamp.

Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher, 43*(6), 304–316.

Mayring, P. (2016). *Einführung in die qualitativ Sozialforschung: Eine Anleitung zu qualitativem Denken* (6. Auflage). Weinheim: Beltz.

Meuser, M., & Nagel, U. (2013). Experteninterviews: Wissenssoziologische Voraussetzungen und methodische Durchführung. In B. Friebertshäuser, A. Langer & A. Prengel (Eds.), *Handbuch qualitative Forschungsmethoden* (pp. 457–471). Weinheim: Beltz.

King, G. (1995). Replication, replication. *Political Science and Politics, 28*(3), 444–452.

Klein, R. A., Vianello, M., Hasselmann, F., Adams, B. G., Adams, R. B., Alper, S., & Nosek, B. A. (2018). Many labs 2: Investigating variation in replicability across

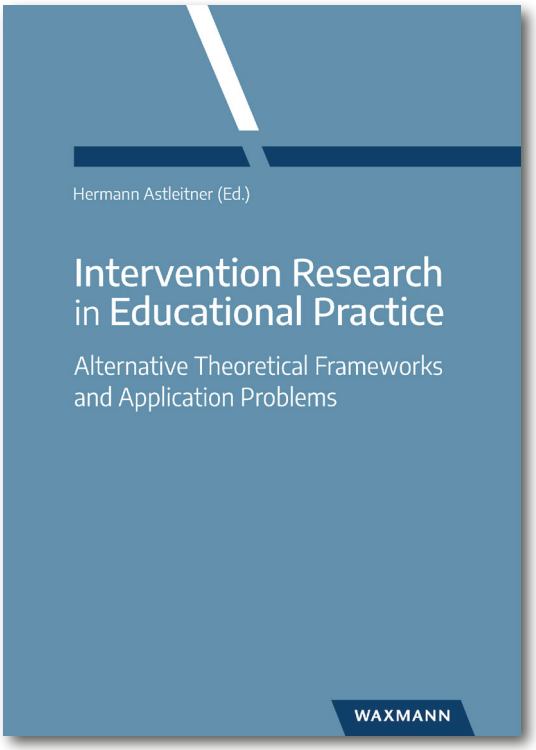
samples and settings. *Advances in Methods and Practices in Psychological Science, 1*, 443–490.
Newby, P. (2013). *Research methods for education*. Hoboken: Taylor and Francis.
Oevermann, U. (2000). Die Methode der Fallrekonstruktion in der Grundlagenforschung sowie der klinischen und pädagogischen Praxis. In K. Krainer (Ed.), *Die Fallrekonstruktion* (pp. 58–157). Frankfurt a. M.: Suhrkamp.
Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251).
Pashler, H., & Wagenmaker, E.-J. (2012). Editors introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science, 7*(6), 528–530.
Popper, K. R. (1935). *Logik der Forschung. Zur Erkenntnistheorie der modernen Naturwissenschaft*. Wien: Springer.
Rost D. H., & Bienefeld M. (2019). Nicht replizieren: publizieren!? *Zeitschrift für Pädagogische Psychologie. 33*(3-4), 163–176.
Schafer, W. D. (2001). Replication: A design principle for field research. *Practical Assessment, Research & Evaluation, 7*(15).
Schmidt, S. (2012). Muss man alles wiederholen? Eine kritische Analyse des Replikationsbegriffs in der modernen Wissenschaft. In W. Ambach (Ed.), *Experimentelle Psychologie in Grenzgebieten*, Vol. 10 (pp. 233–261). Würzburg: Ergon.
Schmidt, S. (2017). Replication. In M. C. Makel & J. A. Plucker (Eds.), *Towards a more perfect psychology: Improving trust, accuracy, and transparency in research* (pp. 233–253). Washington, DC: American Psychological Association.
Sidman, M. (1960). *Tactics of scientific research. Evaluating experimental data in psychology*. New York: Basic Book.
Weber, M. (1904). Die "Objektivität" sozialwissenschaftlicher und sozialpolitischer Erkenntnis. *Archiv für Sozialwissenschaft und Sozialpolitik, 19*(1), 22–87.

Hermann Astleitner (Ed.)

Intervention Research in Educational Practice

Alternative Theoretical Frameworks and Application Problems

2020, 188 pages, pb, € 29,90,
ISBN 978-3-8309-4197-2
E-Book: € 26,99
ISBN 978-3-8309-9197-7



With contributions by
Hermann Astleitner, Josef Eisner, Ulrike Greiner, Franz Hofmann, Michaela Katstaller, Nomy Dickman, Gabriela Gniewosz, Stephanie Musow, Alexander Naumann

Improving learning and teaching at schools or universities may start with choosing evidence-based interventions and practices, but does not end there. To ensure sustainable changes to programs in educational practice, interventions need to address complex issues related to theories, research designs, and measurements. This book presents typical but often overlooked problems in intervention research in educational practice. These problems are embedded in various educational areas such as, amongst others, school effectiveness, instructional design or motivational aspects of teacher trainings.

WAXMANN
www.waxmann.com
info@waxmann.com