4

Qualitative Research – Does it work?

A Discussion of Qualitative Educational Studies and Generation of Evidence

Berit H. Johnsen

Introduction

How is it possible for a qualitative study to generate evidence-based knowledge about inclusive practices in classroom settings? This is a prominent question in current research-methodological discourse in general and a matter of disagreement within the psychological, educational and related sciences. There is therefore good reason to discuss the question in connection with the joint research project *International Classroom Studies of Inclusive Practices – Comparing teaching-learning processes* [26](Johnsen, 2013; Johnsen et al, 2020; WB 04/06, 2006) with participating research teams from the Universities of Belgrade, Ljubljana, Sarajevo, Skopje, Tuzla, Zagreb and Oslo. In order to place the discussion within the context of this project, the article begins with a brief summary of the research.

Inclusive Practices consists of seven classroom studies with the common issue:

> How does school teach in accordance with the pupils' different levels of mastery and needs for support in the learning process (recourses, barriers and dilemmas)?

The studies have different research foci located within a joint didactic-curricular approach consisting of seven main aspects of teaching – learning – developmental processes on micro level (Alexander, 2009) focusing on the individual

26 Hereafter the project (Johnsen, 2013) is referred to as *Inclusive Practices* and Alexander (2000) *Five Cultures*.

Citation of this chapter: Johnsen, B.H. (2020) Qualitative Research – Does it work?. B. H. Johnsen (Ed.), *International classroom studies of inclusive practises (pp.59-77/pp.57-75 in print edition)*. Cappelen Damm Akademisk. https://doi.org./10.23865/noasp.122

pupil in the community of the class: knowledge of the pupil/s – assessment educational intentions – educational content – methods and organisation – communication – care. The seven aspects are embraced by the eighth main aspect of context, which connects the classroom studies with their different contexts on macro level, thus facilitating conditions for comparison (Alexander, 2000; 2004; 2009; Johnsen, 2013; 2014b). The international classroom research is characterized by methodological flexibility within the common denominators of the joint project plan as described in Johnsen's (2014a) summary of *Methodological Diversity in Common Explorations*. The article documents the research teams' application of mostly qualitative or mixed-methods approaches. This actualises the question of whether qualitative research contributes to evidence-based practice. As indicated, the seven participating universities are situated in six countries "on the south-eastern and north-western outskirts of Europe". The joint project is based on a common international ethical-political idea or principle about educational inclusion (Johnsen, 2013; UNESCO, 1994). The goal is to explore the development of inclusive practices, focusing on resources and potential dilemmas and obstacles. As indicated, the university teams have a high degree of freedom concerning their a) research methodology; and b) choice of research focus concerning the elementary school. Thus, the purpose of this article is to discuss the question: Qualitative Research – Does it work?" in light of ongoing "evidence-debates", using the abovementioned project as an example.

How is it possible for a study to generate evidence-based knowledge about "what works" in education?

Why is this a timely question? And, what is meant by evidence-based knowledge or practice? Answering the first question, currently, there is a rising awareness and desire amongst politicians and civil servants to base professional practice on research evidence as a way of ensuring efficient services. Thus, the *Norwegian Directorate for Children, Youth and Family Affairs* (Bufdir) emphasizes that recommended programmes and approaches be evidence-based. Amongst the many links on evidence-based practices found on their homepage is the report from the UK Department for Education (Wiggins, Austberry & Ward, 2012) *Implementing Evidence-Based Programmes in Children's Services: Key Issues for Success*. Using classic snowball method, another similar major report focusing on child welfare services is found, namely UK politician Graham Allen's (2011)

independent report to Her Majesty's Government on *Early Intervention: The Next Steps*. The two reports strongly indicate that focusing on evidence-based knowledge is an international trend – at least in Western societies. Turning to texts within the field of education, evidence-based knowledge is at the centre of a number of anthologies, conference presentations and articles written by educational administrators, stakeholder groups and researchers. In the USA, the U. S. Department of Education follows this trend (Gall, Gall & Borg, 2007). In order to help practitioners assess whether a research project fulfils the criteria needed to be accepted as an evidence-based guide to educational practice, they have published the compendium entitled *Identifying and Implementing Educational Practices Supported by Rigorous Evidence: A User Friendly Guide* (U. S. Department of Education, 2003).

What, then, is meant by evidence-based knowledge or -practice? A fast and informal web search for relevant texts containing the word "evidence-based" in the title reveals that in a majority of these texts, there is no description or clarification of this central concept. On the contrary, its meaning seems to be taken for granted at the same time as it points to a variety of different connections. However, the U. S. Department of Education's guide (2003), mentioned above, gives a clear-cut and concrete description of "what works". Similarly, Allen's (2011) previously mentioned report presents a set of standards for evidence-based statements developed by a team of experts from renowned institutions on both sides of the Atlantic. Thus, a prominent dimension called evaluation quality is described as:

> "… favouring those Early Interventions that have been evaluated to a very high standard using the most robust evaluation methods, such as randomised controlled trials or quasi-experimental techniques, and ideally summarised in systematic reviews" (Allen, 2011: 69).

This conceptual description delimits the field of applicability to natural-science inspired methodology. Where does that leave research based on qualitative methodology?

Educational philosopher Tone Kvernbekk (2013) presents an overview of the use of the term "evidence-based" in educational research. She starts by pointing out that the terminology is suitable to explain "what works" for both learned and laity. The concept may function as a bridge between funded knowledge in research reports and the conventional wisdom of non-professionals[27]. Kvern-

27 The two concepts of "funded knowledge" and "conventional wisdom" stem from John I. Goodlad's classic book on Curriculum Inquiry (1979).

bekk points to educational scholar David Hargreaves as the architect of the view that "…to gather evidence about what works in what circumstances is the whole point of evidence-based research (1996b in Kvernbekk, 2013: 64). He maintains that educational research should improve teaching outcomes. But, how is it possible to determine these outcomes? Hargreaves argues that outcomes are generally perceived as measurable outputs found through studies based on randomised controlled trials. In this way, he places his scientific point of view within quantitative research methodology; as Allen (2011).

Looking back to my student years at the University of Oslo in the early 1970's, two decades before Hargreaves argued in favour of his "what works" position, the natural science based quantitative approach was *the* privileged methodology at the Department of Education, which was then strongly influenced by American educational science. F. N. Kerlinger's methodological handbook *Foundations of Behavioural Research* (1964) was obligatory reading. His primary focus was on collecting and analysing generalizable data that could serve to explain, predict and control educational processes, as Johanningmeier and Richardson (2008) sum up in their historical work on educational methodology. As another example, James R. Lewellen (1977) is amongst the supporters of Kerlinger's ideas. He is concerned with refining social concepts – such as power, conflict, alienation and socialisation – into measurable constructs in terms of observable properties that can be measured.

How is it possible for a qualitative study to generate evidence-based knowledge about practices in classroom settings?

The summative review above provides scarce opportunities for qualitative studies to be accepted as evidence-based. However, a few years ahead of Lewellen's article (1977), John Martin Rich (1975:329) argues that "… the prevailing model, which we call "scientific behavioural" thinking, is not entirely appropriate for fruitful thinking and research in education". Instead, he offers an alternative approach to research on educational practice, namely an idiographic holistic approach not focusing on "uniformities and regularities of a whole class of objects", as done by quantitative research – but on understanding the individual pupil "as a unique being, rather than a specimen of a class" (Rich, 1975: 330). Rich presents an alternative approach to understanding educational practice, or the teacher-pupil relationship, based on Martin Buber's (1947) humanistic "I-Thou"

philosophy comprised of his quest for an inclusive relationship and comprehension of the pupil as a holistic and complex individual within his or her cultural context (Alasuutari, 2010; Johnsen, 2014b; Rich, 1975). Rich outlines a normative approach to educational research, making use of observation, conversation or open interviews, contextual studies and the researcher' experience of empathetic and caring insight, into individual pupils' observable as well as internal set of "values, aims and aspirations". In this way, he offers an alternative approach to Hargreaves in the search for "what works" in education. In a more recent and frequently cited article, Norman K. Denzin (2009) presents a critical overview of the two traditions. His critique of quantitative evidence tradition may be summarised in three main aspects.

1) He strongly opposes the research-political monopoly that is given to quantitative research tradition following Hargreaves' and Kiplinger's position by several funding agencies, professional associations and journals. Consequently, it is given the power to control the definition of evidence.
2) In line with Rich (1975), Denzin argues that evidence is never morally or ethically neutral, as seems to be the view of many followers of the quantitative evidence tradition.
3) He criticises the narrow and conformative basic principles underlying this scientific tradition in contrast to qualitative and interpretive scientific tradition. Where quantitative tradition focuses on uniformity under one set of quality criteria for evidence, qualitative tradition focuses on flexibility in quality criteria that describes evidence adapted to the variety of methodological traditions. Where the quantitative stand favours evidence fit for prediction, qualitative traditions favour the kind of evidence contributing to understanding, thoroughness and awareness of nuances and connectedness to different interpretations. According to qualitative tradition, evidence and data need interpretations and re-interpretations when applied in relation to different research questions, methods and analysis.

Kvernbekk (2013) discusses different viewpoints concerning what is legitimate evidence-based knowledge and the idea about "what works", from Hargreaves' delimited claim for quantifiable randomised controlled trials, to opponents', such as Gert Biesta's rejection of the applicability of evidence-based knowledge at all. Kvernbekk argues that Hargreaves' stand seriously restricts the content and function of the concept of evidence. The basic meaning of evidence is "… that which supports or justifies views, theories, beliefs – and by extension,

teaching strategies or interventions. This function can be performed not only by data, but also by experience, facts, narratives and other reasons" (Kvernbekk, 2013:70-71). In this instance, while Kvernbekk is in line with Denzin (2009), she supports the argument that it is better for educational practice to be based on evidence – in this broad understanding of the term – than to rely on habits or arbitrary opinions.

As may be observed by the above arguments, Kvernbekk's position is similar to mine. Research findings obtained by qualitative methodology – be it action research, case studies, narratives or related research designs – as well as findings through the use of quantified data or mixed methods, all contribute to generate new knowledge about different aspects of the complex process of teaching, learning and development. These are aspects that must be considered in view of their context and the context of the practicing teacher and special needs educator (Denzin, 2009; Foreman-Peck & Murray, 2008; Griffiths & Macleod, 2008; Webb & Ibarz, 2006). There is no single methodology that can claim ownership of "the truth". On the contrary, what counts as evidence is by no means clear-cut, as it depends upon the intended meaning or epistemological foundation of the research question as well as methodology, analyses and interpretation. Different philosophical approaches shed light on different aspects of a phenomenon through their systematic gathering of evidence. In this way, different methodologies contribute to a multifaceted understanding that may advise politicians and practitioners in their search for high-quality answers to their educational questions (Kvernbekk, 2013; Oancea & Pring, 2008).

The increasing use of qualitative studies is accompanied by desires "to develop the quality of qualitative research". This is important for the main example applied in this article, *Inclusive Practices* (Johnsen, 2013; Johnsen et al, 2020), as well as for all other qualitative studies within education and other sciences. As an example, in addition to the educational sciences, qualitative studies are increasing within medical research, which is a field where quantitative methodology has traditionally held a very strong position (Collingridge & Gannt, 2008). Accordingly, refining qualitative methodology is a topic of a growing number of medical research articles, including the question of evidence. Thus, McBrien (2008) recommends four techniques that contribute to the validity of qualitative studies, namely member checking, peer-debriefing, audit trial and reflectivity. He argues that they all contribute to enhancing the research process' credibility, trustworthiness and rigour as well as its outcome; therefore, they are well suited as criteria for evidence-based qualitative studies. Several scholars

discuss and refine techniques in order to develop "the quality of qualitative research"; amongst them the three outstanding scholars, Denzin (2009), Stake (1995; 2006) and Creswell, currently along with Poth (2018). Lincoln & Guba, also outstanding pioneers, direct attention on two complementary main concepts that bring together a number of aspects answering the question of evidence, namely trustworthiness and authenticity (Creswell & Poth, 2018; Guba, 1981; Johnsen et al, 2020; Lincoln & Guba, 1986; Morrow, 2005; Schwandt, 2007). The importance of their contribution lies in their systematic compilation of concepts, descriptions and examples that contribute to develop a conceptual map specifically dedicated to verifying the quality of qualitative studies. Guba and Lincoln's "map for evaluating the quality of qualitative research" is selected as a main approach in *Inclusive Practices* (Johnsen et al, 2020) because of its a) comprehensive and systematic compilation concerning trustworthiness; as well as its b) daring introduction of authenticity as a possible way to further develop the uniqueness of how qualitative research contributes to illuminating "what works". Others have already used most of the terms in their compilation, and many scholars continue to develop the applicability of terms introduced by the two pioneers. Guba (1981) describes the main traits or criteria for trustworthiness as a) credibility, b) confirmability, c) dependability, and d) transferability, while the main characteristics of authenticity are described as i) fairness, ii) ontological authenticity, iii) educative authenticity, iv) catalytic authenticity, and v) tactical authenticity (Lincoln & Guba, 1986). In the following section, each trait is briefly clarified using examples from *Inclusive Practices* (Johnsen, et al, 2020).

Trustworthiness concerns if and how all aspects of a qualitative study, from research issue to report, contain a holistic and nuanced presentation of the phenomena in focus; whether it is worthy of being trusted as evidence. It consists of four main traits, each comprised of tools or techniques used to judge the trustworthiness of any particular study (Guba, 1981; Johnsen et al, 2020).

Credibility is the most comprehensive of the four aspects. It concerns "the truth value" of a study's phenomenon, or if a study is perceived as "true", or valid, by all of its participants and stakeholders, from researchers to practitioners. Credibility concerns all phases in a study; planning, implementing and writing a research report. A number of techniques, or tools, are used in order to establish credibility in each phase. These include establishing structural corroboration, close collaboration, prolonged engagement, triangulation, thick descriptions, member reflections and testing out the correspondence between the seven sin-

gle studies and joint report (Creswell, 2007; Creswell & Miller, 2000; Gall, Gall & Borg, 2007;Geertz, 1973; Gill, Gill & Roulet, 2018; Guba, 1981; Moon et.al, 2016; Northcote, 2012; Tracy, 2010).

In *Inclusive Practices* credibility is considered in the preparation and planning phases, theoretical foundation and methodological choices. However, the main focus is on the process of analysing and compiling the seven individual studies produced by research teams in six European countries into one joint report. Thus, credibility is evaluated in the following phases of the research process:

1) *The question of credibility of preparation and planning* concerns whether the intended research purpose and construction of the joint research project are perceived as meaningful to all participants (Moon et. al., 2016). This includes what Tracy (2010) points to as a worthy topic, or, if it is perceived as relevant, timely, significant, interesting and useful (Johnsen et al, 2020). Three aspects are accounted for here; a) the preparation phase, b) the joint planning phase and c) the team planning of each individual study. a) The preparation phase takes place in an innovation project between the universities of Sarajevo, Tuzla and Oslo, focusing on individually adapted education and inclusion (Johnsen, 2007; SØE 06/02); b) Planning the international comparative classroom study towards inclusion takes place in the application period for inter-European cooperation (Johnsen, 2013a; WB 04/06, 2006). The project plan contains a joint research question and structure of the shared research based on seven didactic-curricular main aspects that constitutes a joint umbrella, or frame, for studies, comparative analysis and discussions of the inner activity of schooling, otherwise called the internal micro dimension (Alexander, 2000; Johnsen, 2013a). c) Each research team develops their own plan based on an eclectic selection of theory- and research traditions within the joint research frame. Are the seven individual study plans and their relation to the joint research plan perceived as meaningful? The joint final research report strongly indicates that they are.
2) *The credibility evaluation of the main implementation phase*. a) Each team conducts their study; b) Six ambulating workshops are held with international researchers participating in discussions focusing on central methodological and theoretical aspects of the seven studies; b) Visits to participating elementary schools are made. c) Sharing the workshops, joint methodology readings and texts describing and discussing the seven research plans and implementation, strengthen the joint understanding between the research teams.

3) The main assessment of credibility takes place in *when collecting, analysing and compiling the joint international comparative report* in a process consisting of eight steps of compiling and revising drafts through a series of email exchanges, a discussion seminar and continuous member checks, revisions and writings, concluding with joint peer or colleague reviews.

The following tools, or techniques, for judging credibility are applied during this process; a) close cooperation; b) prolonged engagement; c) establishing structural corroboration; d) triangulation; e) member reflections; f) checking correspondence between single studies and joint report; g) audit trial; h) and to a lesser degree external peer review (Creswell & Poth, 2018; Creswell & Miller, 2000; Gall, Gall & Borg, 2007;Geertz, 1973; Gill, Gill & Roulet, 2018; Guba, 1981; Johnsen et al, 2020; Lincoln & Guba, 2007; Moon et.al, 2016; Northcote, 2012; Stake, 1995; 2006; Thomas, 2017; Tracy, 2010). Credibility is by far the most extensive aspect of trustworthiness; indeed, this assertion has been discussed and developed by a considerable number of scholars.

Confirmability focuses attention on whether reported findings answer the research issues or are result of research bias. The question is if it is possible to confirm the truthfulness of the research. One criterion for confirmability is therefore that it must be possible to replicate a similar research process and come to similar conclusions – to the extent that this can be realized in qualitative studies within different contexts. How is it possible to account for possible biases? Four techniques are used in *Inclusive Practices* to account for the different interpretations, operationalisations and choices in order to reveal as clearly as possible the research process: a) Accounting for the underlying assumptions leading to the construction of the research. b) Ensuring that interpretations and conclusions are grounded in evidence; c) Giving detailed methodological descriptions. d) Making use of internal and external auditing (Johnsen et al, 2020).

Dependability applies to the findings' stability and consistency. Qualitative studies are not suitable for direct replications, but accuracy, logical consistency and possibility of an approximately similarly perceived research process are hallmarks of trustworthiness (Anney, 2014; Armstrong, 2010; Gill, Gill & Roulet, 2018; Guba, 1981; Moon et al., 2016). Dependability auditing is a major assessment technique that consists of external peer audits' review of the entire research process. Several methods contribute to assessing dependability, such as a triangulation, stepwise replications, also called "dependability audit", cod-

ing-re-coding strategy; peer examination and audit trial. The dependability of *Inclusive Practices* is confirmed through the eight steps described above as a credibility check that contains collecting, analysing and revising processes of the seven studies based on the shared didactic-curricular main aspects. The stepwise procedure consists of a series of internal audits (Johnsen et al, 2020).

Transferability. While dependability focuses on the research process and findings, transferability mainly concerns whether or not the results can be transferred to other contexts. Geertz' (1973) thick descriptions, where findings are described in their context are therefore the main criteria used to determine truthfulness of results; they are also used for determining credibility, as mentioned above. Transferability has been compared to external validity or the validity of applying a qualitative study's conclusions outside the context of that study. Thick descriptions may involve illuminating all parts of the research process, from background data, phenomenon, research questions and choice of methods, situations, informants and data collection, to findings and compilation of the final report. Hence, thick descriptions based on contextual disclosures contribute to transferable truth-value and pave the way for replicating the study in other settings. The contextual descriptions of the seven studies that make up this international comparative research are therefore crucial for transferability. (Anney, 2014; Gill, Gill & Roulet, 2018; Guba, 1981; Johnsen et al, 2020; Korstjens & Moser, 2018; Moon et al, 2016; Schwandt, 2007; Shenton, 2004).

The authenticity perspective draws attention to a dimension that is unique for ideographic, qualitative research and is characterized by its "… relativist ontology and an interactive, value-bounded epistemology" (Lincoln & Guba, 1986: 20). Hence, authenticity applies to value awareness. In their introduction of the authenticity perspective of methodological rigor, Lincoln and Guba (1986) admit that they have not yet developed this perspective fully, especially when it comes to assessment methods. However, they suggest five aspects of authenticity; fairness, ontological authenticity, educative authenticity, catalytic authenticity, and tactical authenticity. Even though the authenticity quality perspective is not applied in a large number of qualitative studies, several scholars are engaged in further discussions about its development and use (Lincoln & Guba, 1986; Manning, 1997; Schwandt, 2007; Shannon & Hambacher, 2014). In the following, the five criteria of authenticity are briefly described using examples from *Inclusive Practices*.

Fairness: Among the five aspects of authenticity, fairness is considered the most outstanding. It is based on the following line of arguments; i) that qualitative or naturalistic studies are value-based, ii) that they are constructed in accordance with differing value systems, and iii) that an important part of qualitative research is to account for its value structures. Consequently, it is fair 1) that the researcher explicitly discusses the inquiry's value framework and, as Manning (1997) argues, 2) that all participants have a voice in the inquiry. Manning (1997) also presents an extensive list of tools to assess the two aspects of fairness. Several of these assessment tools are also used in assessing trustworthiness. Thus, the same assessment tool, or technique, considers several aspects of research quality, as also occurs in the presentations of trustworthiness. The internal peer debriefing in the 8-step process of the joint research report's compilation of *Inclusive Practices* (Johnsen et al., 2020) is possibly the most prominent example of this, being a quality criterion related to credibility, fairness and several other aspects of "goodness of qualitative research". The two main aspects of fairness consider a) fairness as describing and discussing the research's value framework, and b) that all participants have a voice in the inquiry. In *Inclusive Practices* the underlying value framework is discussed with a focus on; a1) theoretical considerations; a2) international human rights principles; and a3) underlying basic value considerations when focusing on a critical analysis of good examples of educational inclusion; b) all participating researchers have a voice in the compilation and revision process of the abovementioned 8 steps leading up to the joint research report (Johnsen et al, 2020). However, in each of the seven studies, there are different participants and stakeholders whose voices are important and fair and should therefore be accounted for in each of the individual studies.

By applying the four authenticities – ontological-, educative -, catalytic- and tactical authenticity – Lincoln and Guba (1986) shed light on nuances of increased understanding, applicability and societal relevance with respect to qualitative studies (Johnson & Rasulove, 2017; Lincoln and Guba, 1986; Schwandt, 2007; Shannon & Hambacher, 2014). Thus, **ontological authenticity** concerns whether or not participants in a study gain increased experience of the complexity of a phenomenon, such as in *Inclusive Practices*, when they experience a) the significance of the interrelations between the seven didactic main areas of the didactic-curricular relation approach (Johnsen, 2014b) – the pupil/s – educational assessment – educational intentions – educational content – methods and classroom organisation – communication – care – context /frame factors –,

as well as relevant sub-areas, in practicing individually adapted teaching for all pupils in the community of the class; b) and when they apprehend the important role that the close and wider context plays in classroom practices. When participants also become aware that the process of the inquiry has led to their own reconstruction towards gaining an increased understanding of the complexity of the practice mentioned above, as well as different value systems, they have at that point acquired **educative authenticity**.

Lincoln and Guba (1986) argue that applicability is a criterion of good quality. They also point out that studies should facilitate and stimulate action, calling this feedback validity. Assessing **catalytic authenticity** in our research example therefore consists of examining if and how the inquiry process stimulates stakeholders' engagement. **Tactical authenticity** focuses on all participants in a study, which in this example means pupils, parents, teachers, special needs educators, principal and school administration, as also the research teams themselves. In addition to participants, other stakeholders should also be mentioned, such as local and national politicians and officials as well as higher education institutions and researchers within the fields of education and special needs education. The criteria of tactical authenticity include if the findings are empowering or impoverishing for the different participants and interest groups involved in the research project (Johnsen et al, 2020).

Does the research project *International Comparative Classroom Studies towards Inclusion* (Johnsen et al, 2020) generate evidence-based knowledge about practices in classroom settings? Several or tools or techniques are used to examine the quality and "truth value" of this qualitative research project – in other words, its strengths and limitations in generating evidence. They indicate that of the many positive quality checks, the most prominent strength is the combination of close collaboration in prolonged engagement together with systematic, repeated internal auditing, multivocality and reflections. When taken together, they strongly indicate that the research contains truthfulness and authenticity. The weakest link is the limited and unsystematic external audits, as they represent a limitation of the quality check.

An international comparative qualitative research project

Do arguments for qualitative evidence generation also apply to comparative studies? Kvernbekk's argument that if only research findings obtained by quan-

tifiable randomised controlled trials are accepted, this would seriously restrict the content and function of the concept of evidence, is supported by Robin Alexander, main editor of the extensive *Cambridge Primary Review Research Surv*eys (2010). He is also author of *Culture and Pedagogy* (2000), an extensive research report on international comparisons of primary education in five countries on three continents. Alexander asks: "In comparing ourselves with others have we got the balance of evidence right? Are we taking too much notice of some kinds of evidence and too little of others?" (Alexander, 2012: 3). He warns: "The way the discourse of international comparison is dominated by international achievement surveys and the accompanying media and political hysteria requires us to think more deeply about evidence" (Alexander, 2012: 4). Referring to an official and highly regarded report, he argues the following (Alexander, 2012: 4):

> What the … report is saying, if we can express the matter even more bluntly, is that in pursuit of what they call 'evidence-based policy', governments choose to ignore the larger part of the international evidence that is available to them, including evidence that could give them the insights, explanations and policy options they need.

What is the essence of Alexander's critique? His main concern is that school advocates, media and politicians choose to seek knowledge from an aspect of the complexity of pedagogical practice that is too limited. He criticises 1) the narrow empirical arena when it concentrates solely on students' learning outcomes; 2) the preferred focus on quantitative studies; and 3) the consequence of accepting the narrow application of the concept of evidence, which results in there only being a limited part of school-related research that is accepted by media as well as politicians. As a result, Alexander argues in favour of a broad application of the evidence concept. Focusing on comparative studies in particular, he points out that a broad evidence concept includes the majority of studies in the published corpus of academic international and comparative education. These range from descriptive accounts of individual education systems to in-depth studies of school and classroom life related to their historical and sociocultural contexts (Alexander, 2012). Applied to studies of teaching-learning processes in general, it seems that a broad application of the evidence concept invites politicians as well as public officials and media to be aware of a much larger and more nuanced corpus of research – while the evidence concept continues to enjoy its privileged position as a quality mark.

Conclusion

Is it possible for a qualitative study to generate evidence-based knowledge about inclusive practices in classroom settings? And if so, how is it possible? These questions, which were posed at the beginning of this article, are currently important due to the previously mentioned rising popularity of the concept of evidence-based knowledge within both public debate and scientific communities. The questions are of specific relevance for the joint research project *International Comparative Classroom Studies towards Inclusion* (Johnsen, 2013; Johnsen et al, 2020). Investigation of what is meant by evidence-based knowledge shows two main trends – narrow and broad – either reflecting whether advocates of the importance of evidence-based knowledge are referring to a delimited or the entire research universe. The narrow trend has been recommended by representatives situated in the same methodological discourse and is widely accepted by nonprofessional society. What actually characterises these two trends?

The narrow trend is based on the view that only research findings obtained by quantifiable randomised controlled trials are acceptable as evidence-based knowledge that is suitable for application in educational practice. This view has many advocates within the research community, from the architect of the so-called "what-works" statement David Hargreaves and his likeminded colleagues, such as Kerlinger, Lewellen and Allen, to the U. S. Department of Education's adoption of this kind of criteria for WWC – or what works in education (U.S. Department of Education, 2003).

The broad application of the evidence concept emphasises the following:

a) That there is no one methodology that can claim ownership of "the truth"; knowledge about a phenomenon depends upon the construction and epistemological basic of the research issue as well as other methodological aspects
b) That to recognise the majority of research presentations in the published corpus of educational studies contributes to broad, in-depth and nuanced knowledge about educational practices
c) That it is of specific importance in educational inclusion to understand the pupil as a unique being rather than a specimen of a class since this is fundamental to understanding the complex teaching-learning processes within the diversity of a school class (Alexander, 2012; Buber, 1947; Creswell & Poth, 2018; Denzin, 2009; Johnsen, 2014b; Johnsen et al, 2020; Johnson & Rasulova, 2017; Rich, 1975).

When exploring the literature, it may seem that the narrow application of the concept of evidence-based knowledge has a stronger position than the broad application. The U.S. Department of Education's placement within this narrow trend may contribute to its privileged position when it comes to research funding as well as the application of research findings. It is, however, interesting to observe how the classical introductory handbook, Educational Research (Gall, Gall & Borg, 2007) has revised and extended its discussion of what the authors call "the nature of educational research". It must be mentioned that they do not argue against the U.S. Department of Education's narrow understanding of research evidence. Nonetheless, through the steady publication of new editions of this internationally read methodology book, they have given increasing emphasis to qualitative methodologies by introducing new ones in new editions. Together with the increasing variety of methodologies, designs and methods, a theory of science discussion is currently growing, which is related to how research focus and methodology are chosen. Last but not least, a series of important research quality criteria reaching far beyond the abovementioned narrow evidence criteria is in development, whereof the urge to minimize research errors and biases is of basic importance for all kinds of research.

Where is the research project *International Comparative Classroom Studies towards Inclusion* (Johnsen et al, 2020) situated in the dispute between these narrow and broad understandings of evidence? Consisting of seven independent research projects within a joint – yet flexible – frame, located in different cultures, applying qualitative methodologies, and mixed methods, the research project is situated within the broad application of the evidence-based research term. The research project is also, as its title implies, an international comparative classroom study having the intention of critically exploring and finding evidence of inclusive practices as well as its dilemmas and challenges. Several scholars within international comparative educational studies support a broad understanding of this field (Johnsen, 2020). Robin Alexander's major international comparative work, *Culture and Pedagogy* (2000), is an example of the search for evidence across cultures and research methodologies. Both his work and arguments (Alexander, 2000; 2010; 2012) have been of relevance and importance for planning and implementing *Inclusive Practices*.

Finally, "does the research project work"? Has it contributed to knowledge about and critical reflections on inclusive practices? The answer to this question may be found in the report that both presents and discusses the overall findings in this exploratory research project. It may also be found in the discussions of

dilemmas and challenges in the process of developing educational inclusion, as well as in a more detailed examination of the methodological criteria used in evidence-based qualitative research (Johnsen et. al., 2020).

References

Alasuutari, P. (2010). The Rise and Relevance of Qualitative Research. In *International Journal of Social Research Methodology*, 13:2, 139–155, DOI:10.1080/13645570902966056.

Alexander, R. (2012). *International Evidence, National Policy and Educational Practice. Making Better Use of International Comparisons in Education*. Keynote at the July 2012 international teacher education conference at Canterbury Christ Church University. http://www.robinalexander.org.uk/wp-content/uploads/2014/05/alexander-Jerusalem_Canterbury.pdf.

Alexander, R. (Ed.). (2010). *The Cambridge Primary Review Research Surveys*. London: Routledge.

Alexander, R. (2009). Towards a Comparative Pedagogy. In R. Cowen & A. M. Kazamias (Eds.). *International Handbook of Comparative Education*, Part 2 (pp. 923–942). Dordrecht: Springer.

Alexander, R. (2004). Still No Pedagogy? Principle, Pragmatism and Compliance in primary Education. In *Cambridge Journal of Education*, 34: 1, 7–33. DOI: 10.1080/0305764042000183106.

Alexander, R. (2000). *Culture & Pedagogy. International Comparison in Primary Education*. Oxford: Blackwel Publishing.

Allen, G. (2011). Independent Report to Her Majesty's Government on Early Intervention: *The Next Steps* Independent Report to Her Majesty's Government.

Anney, V. N. (2014). Ensuring the Quality of the Findings of Qualitative Research. Looking at Trustworthiness Criteria. In *Journal of Emerging Trends in Educational Research and Policy Studies* (JETERAPS) 5(2): 272–281. jeteraps.scholarlinkresearch.org.

Armstrong, J. (2010). Naturalistic Inquiry. In N. J. Salkind (Ed.). *Encyclopaedia of Research Design* (pp. 881–885).Thousand Oaks: SAGE Publications, Inc. http://dx.doi.org/10.4135/9781412961288.n262.

Buber, M. (1947). *Between Man and Man*. Great Brittain: Kegan Paul, Trench, Trubner & Co. Ltd.

Collingridge, D. S & Gantt, E. E. (2008). The Quality of Qualitative Research. In *American Journal of Medical Quality*, vol. 23, 5: pp. 389–395. Doi: 10.1177/1062860608320646.

Creswell, J. W. & Poth, C. N. (2018). *Qualitative Inquiry and Research Design*. Thousand Oaks: Sage Publications.

Creswell, J. W. & Miller, D. L. (2000). Determining Validity in Qualitative Inquiry. In *Theory into Practice*, 39:3, 124–130, DOI: 1207/s 15430421tip3903 2.

Denzin, N. K. (2009). The Elephant in the Living Room: on Extending the Conversation about the Politics of Evidence. In *Qualitative Research*, 9:2, 139–160. http://qrj.sagepub.com/.

Foreman-Peck, L. & Murray, J. (2008). Action Research and Policy. In *Journal of Philosophy of Education*, Vol. 42: S1, 145–163.

Gall, M. D., Gall, J. P. & Borg, W. R. (2007). *Educational Research*. Boston: Pearson Education, Inc.

Geertz, C. (1973). Deep Play: Notes on the Balinese Cockfight. In C. Geertz (Ed.). *The Interpretation of Cultures* (pp. 412–454). New York: Basic Books, Inc., Publishers.

Gill, M. J., Gill, D. J. & Roulet, T. J. (2018). Constructing Trustworthy Historical Narratives: Criteria, Principles and Techniques. In *British Journal of Management*, Vol. 29, 191–205. DOI: 10.1111/1467-8551.12262. https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-8551.12262.

Goodlad, J. I. (1979). *Curriculum Inquiry*. New York: McGraw-Hill Book Company.

Griffiths, M. & Macleod, G. (2008). Personal Narratives and Policy: Never the Twain? In *Journal of Philosophy of Education*, Vol. 42: S1, 121–143.

Guba, E. G. (1981). Criteria for Assessing the Trustworthiness of Naturalistic Inquiries. In Educational Communication and Technology Journal, 29, 2, 75–91.

Johanningmeier, E. V. & Richardson, T. (2008). *Educational Research, the National Agenda, and Educational Reform: A History*. USA: IAP-Information Age Publishing, INC.

Johnsen, B. H. (2020). Comparing Classroom Activities – International Comparison in Qualitative Pedagogical Studies. In B. H. Johnsen (Ed.). *International Classroom Studies of Inclusive Practices – Comparing Teaching-Learning Processes*. Oslo: Cappelen Damm Akademisk.

Johnsen, B. H. (2014a). Methodological Diversity in Common Explorations. Seven Research Communities Collaborating in International Comparative Classroom Studies towards Inclusion. In B. H. Johnsen (Ed.). *Theory and Methodology in International Comparative Classroom Studies* (pp. 267–275). Oslo: Cappelen Damm Akademisk.

Johnsen, B. H. (2014b). A Curricular Approach to Inclusive Education. Some Thoughts concerning Practice, Innovation and Research. In B. H. Johnsen (Ed.). *Theory and Methodology in International Comparative Classroom Studies* (pp. 133–183). Oslo: Cappelen Damm Akademisk.

Johnsen, B. H. (2013a). Comparative Classroom Studies towards Inclusion. Joint Research Plan for Cooperation between the Universities of Belgrade, Ljubljana, Sarajevo, Skopje, Tuzla, Zagreb and Oslo. In B. H. Johnsen (Ed.). *Research Project Preparation within Education and Special Needs Education* (pp. 228–244). Oslo: Cappelen Damm Akademisk.

Johnsen, B. H., Ljiljana Igrić, L., Rapaić, D., Salihović, N., Jachova, Z., Košir, S., Zečić, S., Džemidžić Kristiansen, S., Angeloska-Galevska, N., Cvitković, D., Dizdarević, A., Karovska, A., Galevska, Hadžić, S., Ilic, S., Kogovšek, D., Lisak, N., Nedovic, G., Ozbič, M. & Stojkovic, I. (2020). International Classroom Studies of Inclusive Practices. Comparing teaching-learning processes. In B. H. Johnsen (Ed.). *International Classroom Studies of Inclusive Practices – Comparing Teaching-Learning Processes*. Oslo: Cappelen Damm Akademisk.

Johnson, S. & Rasulova, S. (2017). Qualitative Research and the Evaluation of Development Impact: Incorporating Authenticity into the Assessment of Rigour. In *Journal of Development Effectiveness*, 9:2, 263–276, DOI: 10.1080/19439342.2017.1306577.

Kerlinger, F. N. (1964). *Foundations of Behavioural Research*. Holt, Reinhart and Winston.

Korstjens, I. & Moser, A. (2018). Series: Practical Guidance to Qualitative Research. Part 4: Trustworthiness and Publishing. In *European Journal of General Practice*, 24:1, 120–124, DOI:10.1080/13814788.2017.1375092.

Kvernbekk, T. (2013). Evidence-Based Practice and Educational Research. In B. H. Johnsen (Ed.). *Research Project Preparation within Education and Special Needs Education* (pp. 60–72). Oslo: Cappelen Damm Akademisk.

Lewellen, J. R. (1977). The Process of Social Inquiry. In *The Educational Forum*, 41:2, 199–205, DOI: 10.1080/00131727709336234.

Lincoln, Y. S. & Guba, E. G. (1986/2007). But Is It Rigorous? Trustworthiness and Authenticity in Naturalistic Evaluation. In *New Directions for Evaluation*, special issue 114, 15–25. www.interscience.wiley.com. DOI:10.1002/ev.223.

Manning, K. (1997). Authenticity in Constructivist Inquiry: Methodological Considerations without Prescription. In *Qualitative Inquiry*, 3(1), 93–116. https://doi.org/10.1177/107780049700300105.

McBrien, B. (2008). Evidence-Based Care: enhancing the rigour of a qualitative study. In *British Journal of Nursing*, Vol 17, No 20. DOI: 10.12968/bjon.2008.17.20.31645.

Moon, K., Brewer, T. D., Januchowski-Hartley, S. R., Adams, V. M. & Blackman, D. A. (2016). A Guideline to Improve Qualitative Social Science Publishing in Ecology and Conservation Journals. *Ecology and Society*, 21(3):17. http://dx.doi.org/10.5751/ES-08663-210317.

Morrow, S. L. (2005). Quality and Trustworthiness in Qualitative Research in Counselling Psychology. In *Journal of Counselling Psychology*, 52, 2, 250–260. DOI: 10.1037/0022-0167.52.2.250.

Northcote, M.T. (2012). "Selecting Criteria to Evaluate Qualitative Research" (2012). *Education Papers and Journal Articles*. Paper 38. http://research.avondale.edu.au/edu_papers/38.

Norwegian Directorate for Children, Youth and Family Affairs (Bufetat). https://www.regjeringen.no/en/dep/bfd/id298/.

Oancea, A. & Pring, R. (2008). The Importance of Being Thorough: On Systematic Accumulations of 'What Works' in Education Research. In *Journal of Philosophy of Education*, Vol. 42: S1, 15–39.

Rich, J. M. (1975). Idiographic Holistic Thinking in Education. In *The Educational Forum*, 39:3, 329–333, DOI: 10.1080/00131727509339076.

Schwandt, T. A. (2007). Judging Interpretations. In *New Directions for Evaluation*, no. 114: 11–14. www.interscience.wiley.com. DOI: 10.1002/ev.223.

Shannon, P. & Hambacher, E. (2014). Authenticity in Constructivist Inquiry: Assessing an Elusive Construct. *The Qualitative Report*, 19(52), 1–13. https://nsuworks.nova.edu/tqr/vol19/iss52/3.

Shenton, A. K. (2004). Strategies for Ensuring Trustworthiness in Qualitative Research Projects. In *Education for Information* 22(2):63–75. DOI: 10.3233/EFI-2004-22201.

Stake, R. E. (2006). *Multiple Case Study Analysis*. New York: The Guilford Press.

Stake, R. E. (1995). *The Art of Case Study Research*. Thousand Oaks: Sage Publications.

SØE 06/02. (2002). *Institutional Competence Building and Cooperation with Two Bosnian Universities: "Special Needs Education towards Inclusion"*. The Cooperation Programme with South-East Europe (CPSEE) 2002–2004.

Thomas, D. R. (2017). Feedback from Research Participants: are Member Checks Useful in Qualitative Research? In *Qualitative Research in Psychology*, 14:1, 23–41, DOI:10.1080/14780887.2016.1219435.

Tracy, S. J. (2010). Qualitative quality: Eight Big–Tent criteria for excellent qualitative research. In *Qualitative Inquiry*, 16(10), 837–851.

UNESCO. (1994). *The Salamanca Statement and Framework for Action on Special Needs Education*. Paris: (The author).

U.S. Department of Education. (2003). Identifying and Implementing Educational Practices Supported by Rigorous Evidence: A User Friendly Guide. Washington, DC: The author. www.excelgov.org/evidence.

WB 04/06. (2006). *Development towards the Inclusive School: Practices – Research – Capacity Building: Universities of Belgrade, Ljubljana, Sarajevo, Skopje, Tuzla, Zagreb & Oslo*.

Webb, S. & Ibarz, T. (2006). Reflecting on 'What Works' in Education Research-Policy/Practice Relationships. In *Journal of Vocational Education and Training*, 58:2, 205–222, DOI: 10.1080/13636820600811943.