



Technische Universität München

TUM School of Education

Educational Research in the Mirror of Nature

Theoretical, Epistemological, and Empirical Aspects of Mixed-Method Approaches
in Outdoor Education

Ulrich Dettweiler

Vollständiger Abdruck der von der Fakultät TUM School of Education der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Philosophie genehmigten Dissertation.

Vorsitzende: Univ.-Prof. Dr. Doris Lewalter

Prüfer der Dissertation: 1. Univ.-Prof. Dr. Claudia Kugelman (i.R.)

2. Priv.-Doz. Dr. Jörg Wernecke

3. Univ.-Prof. Dr. Ali Ünlü

Die Dissertation wurde am 24.02.2015 bei der Technischen Universität München eingereicht und durch die Fakultät TUM School of Education am 12.05.2015 angenommen.

1. FROM THE SEA TO THE MOUNTAINS: INTRODUCTION AND THEORETICAL SCOPE.....	5
1.1. Preliminary remarks	5
1.2. Facts and values: educational research in the mirror of nature	8
1.3. A brief history of outdoor education research.....	11
2. OUTDOOR EDUCATION RESEARCH: MERGING THE TWO EPISTEMOLOGICAL PARADIGMS INTO MIXED-METHOD APPROACHES	16
2.1. The so-called “quantitative” approach: measuring “facts”	16
2.1.1. Measurement and scales	16
2.1.2. Ontological presumptions of the stochastic world of facts.....	21
2.2. Measuring “facts” in outdoor education programmes: Janne Fengler’s evaluative study on self-concept.....	23
2.2.1. Research topic, research goal, and research problem.....	23
2.2.2. Status of the research and its perspective	24
2.2.3. Theoretical basis and model use.....	24
2.2.4. Methodology that has been applied.....	27
2.2.5. Epistemological considerations	32
2.3. The so-called “qualitative” approach: aiming at understanding meaning.....	39
2.3.1. The significance of scales.....	39
2.3.2. Tying up the ends of theories of understanding	42
2.3.3. Holistic semantics, interpretation and abductive inferentialism.....	43
2.3.4. Ontological presumptions of the world of meaning(s) and values.....	45
2.3.5. Leaving the methodological dualism behind: Approaching the “Third Paradigm”.....	49
3. MIXED-METHOD RESEARCH IN THE MIRROR OF NATURE: TWO EMPIRICAL SURVEYS WITHIN THE “THIRD PARADIGM”.....	53
3.1. Alien at home... ..	53
3.1.1. Summary, research problem and theoretical model.....	53
3.1.2. Methods of data collection and empirical model.....	53
3.1.3. Methods of data interpretation and statistical model.....	54
3.1.4. Findings and Epistemological considerations	56
3.2. Pupils’ learning motivation during outdoor science teaching.....	57
3.2.1. Summary, research problem and theoretical model.....	57
3.2.2. Methods of data collection and empirical model.....	59
3.2.3. Methods of data interpretation and statistical model.....	60
3.2.4. Findings and Epistemological considerations	61
4. CONCLUSION AND ACKNOWLEDGEMENTS.....	65
5. APPENDIX.....	68
6. BIBLIOGRAPHY	69

1. From the Sea to the Mountains: Introduction and Theoretical Scope

1.1. Preliminary remarks

The three essays attached to this dissertation mark the thematic scope and the historical development of outdoor educational research at Technische Universität München (TUM) since 2010, when Dr. Claudia Kugelmann became the chair of “Sportpädagogik” at the Faculty for Health and Sports Science.

Previously, she had steered the project “KUS – Klassenzimmer unter Segeln” at Friedrich-Alexander Universität Erlangen-Nürnberg, where 32 boys and girls embark the traditional sailing vessel Thor Heyerdahl for a six months over-sea’s educational programme every year. With KUS, the youths follow the paths of great explorers like Alexander von Humboldt and Christopher Columbus and sail from Germany via the Canary Islands into the New World. There, the students explore countries of the Caribbean and Americas during several weeks of layover. The voyage home leads them back to Germany via the Bermudas and Azores. The pupils responsibly participate in ship operations, explore foreign countries and cultures, and are taught on shore as well as at sea a condensed 10th grade school curriculum. The aim of the programme is to strengthen the personality of students. The voyage requires mastering new life situations that demand autonomous decision-making. The exceptional framework conditions of the project provide first-hand experience and adventure. The object of KUS is to strengthen young individuals’ autonomy, initiative taking, and sense of responsibility, as well as to prepare them for the demands of a complex and globalised world (Dettweiler & Kugelmann, 2010).

Despite KUS remaining formally at FAU, two researchers at TUM, Boris Wolter and Gabriele Lauterbach, were involved with the project, basically with summing up the project (Kugelmann & Lauterbach, 2011). My role was to assist Boris and Gabriele in their work, which resulted in a presentation at the annual conference of the “Kommission Sportpädagogik” of the “Deutsche Gesellschaft für Erziehungswissenschaft” in December 2010 (Dettweiler & Wolter, 2010). Shortly after, Boris left TUM and I took over the evaluation work with the KUS-project, together with Gabriele, and also advanced the development of an Alpine equivalent to KUS.

For the latter, I made contact with Dr. Michael Streifinger from the geography-department at Ludwig Maximilians Universität München (LMU), who had just completed his doctoral thesis developing and examining an Alpine excursion project (Streifinger, 2010). I then invited him to progress the Würzburg-presentation into a publication. And this essay became our pedagogical “manifesto” – with the inclusion of Dr. Claudia Kugelmann as a second co-author. The paper is attached to the appendix and is part of the dissertation. It describes the philosophical and educational

background of our understanding of outdoor education in the sense of expeditionary learning, and the didactical implication of using expeditions as an educational tool (Dettweiler, Kugelmann, & Streifinger, 2011).

The concept merges two educational theories – the German “Bildungstheorie” in the tradition of the “Critical Theory” of the Frankfurt-School (Adorno, 2006; Becker, 2005; Bietz, Laging, & Roscher, 2005; Klafki, 2005, 2007; Klafki & Braun, 2007; Liessmann, 2006) and the American tradition of “Outward Bound Expeditionary Learning” (Bushweller, 1997; Campbell et al., 1996; Cousins, 1999). The central concept of the German “Bildungstheorie” is that education/literacy (“Bildung”) is an embodied (“leiblich”; (Böhme, 2002, 2003)) and aesthetic process of interaction of an individual with the world. It encompasses didactic guidelines as situational place-based, and every-day-life relevance, person centred orientation, gender justice, and experiential learning. The corresponding didactical methods include “learning against resistance”, “learning by doing”, group and project work, individualization of teaching processes, or “learning by teaching”.

Parallel to approaching the Alps didactically, I analysed the existing data sets of the KUS-project together with Gabriele, and we decided to set up a completely new approach – since the previous attempts seemed rather low in quality. However, the only possibility was to access the KUS-participants (“KUSis”) ex-post. So I decided to look for adjustment symptoms after their return from the cruise. Herby, I was very much inspired by an article by Dr. Pete Allison, a friend and colleague from the University of Edinburgh (Allison, Davis-Berman, & Berman, 2011) and a discussion with Dr. Bob Henderson from McMaster University, Hamilton/Ontario, who had invited me to contribute to “Pathways: The Ontario Journal of Outdoor Education” with an essay about adjustment symptoms of professional expedition leaders after their return (Dettweiler, 2012).

So, we conducted a postal survey (N=128, global response quote $r_g = .44$) with pupils who had returned from four separate learning expeditions from 2008-2012, asking them to tell us about their experiences of re-adjustment. The pupils’ reports were then coded independently by Gabriele Lauterbach and Andrea Legl, and together we came up with five categories in a consensus decision, i.e. (1) reintegration narratives (RN), (2) perception of schooling (PoS), (3) self-perception (SP), (4) perceived programme effects (PPE), and (5) social context (SC), which were also quantified on a 5-point Likert-scale. I submitted a first draft of the paper to the International Journal of Intercultural Relations, and the anonymous reviewer asked me to tie the findings more to the UU-curve theory of reverse culture shock. With the help of Dr. Ali Ünlü, professor of methods in empirical educational research, I modelled the data with respect to the UU-curve theory and searched for underlying distribution patterns in the categories by means of principal components analyses (PCA) and polynomial interpolation (PI). We found no significant differences with respect to gender in the four groups in a one-way ANOVA. However, significant differences with respect to self-perception (SP) and perceived programme effects (PPE) can be referred to those who had been at home for one year (group 3). This accords well with the findings of the PCA and PI, which corroborate a U-curve with its apex after eight months at home. We thus demonstrated that all four cohorts show symptoms of expedition reverse

culture shock. This indicates that pedagogical intervention also after the cruise may be undertaken.

After revision, this paper has been accepted at the *International Journal of Intercultural Relations* (Dettweiler, Ünlü, Lauterbach, Legl, et al., 2015) – and it marks our approach to empirical research with mixed-methodology, and is the second paper of the dissertation attached in the appendix.

Simultaneously to working on the KUS-ERCS-paper, I developed a proto-type of an expeditionary learning course in the Alps and created the ELPIN-project. At this time, Dr. Andreas Kratzer approached me and asked me to deliver an outdoor educational programme at the newly founded Schülerforschungszentrum Berchtesgaden Land (SFZ-BGL). He also made contact with Dr. Hans Kern, vice-headmaster of Schyregymnasium in Pfaffenhofen, and the three of us decided to try out the ELPIN-course at the SFZ-BGL. I included Dr. Susanne Bley and Dr. Gert Helms from the SFZ-BGL in the development of the contents of the pupils' field-work, and asked one of my students, Christoph Becker, if he wanted to write his diploma thesis evaluating the project. A 25.000 Euro grant from the Universität Bayern e.V. that I had applied for together with 10.000 Euro from Kiel&Ruder e.V. made this project possible.

The pilot-programme in summer 2012 was, despite its huge logistical challenge, a great success, so we decided to go on with the project and I received another 15.000 Euro from Hans Sauer Stiftung for including more pupils in 2013. These became the subjects of the evaluation paper “Investigating the motivational behaviour of pupils during outdoor science teaching within self-determination theory”, published in *Frontiers in Educational Psychology* (Dettweiler, Ünlü, Lauterbach, Becker, & Gschrey, 2015). This is the third paper of this dissertation, attached in the appendix. It may be noteworthy, that the ELPIN research weeks have been awarded the Grand Prize at the “School meets Science” challenge by the Robert Bosch Stiftung in 2013, with a prize money of 50.000 Euro.

For our evaluative research, we again used a mixed-methods approach to search into learning psychological aspects of the ELPIN project. We used data from qualitative explorations into the pupils' learning motivation during field observation, a group interview, and open questionnaires, in order to understand quantitative measures from the Self-Determination Index (SDI) of the German version of the Academic Self-Regulation Questionnaire (SRQ-A) originally developed by Ryan and Connell (1989) and adopted especially for our target group by Müller, Hanfstingl, and Andreitz (2007), and the Practical Orientation (PO) of the programme. For the latter, we used an adopted inventory originally developed by Rakoczy, Buff, and Lipowsky (2005) and validated it especially for our target group.

Our data suggest that lower self-regulated pupils in “normal” science classes show a significantly higher self-regulated learning motivational behaviour in the outdoor educational setting ($p < 10^{-4}$), and that the outdoor-teaching has generally been perceived as more practical than teaching at the normal school context ($p < 10^{-4}$), irrespective of gender or school culture. We provided in-depth analyses of all quantitative findings with our qualitative data and can thus explain the findings logically,

with respect to the direction of the statistical interpretation, and substantially, with respect to the meaning of the discoveries. We conclude that expeditionary outdoor programming appears to be a suitable tool to trigger interest in science in pupils, especially for less motivated ones.

The “journey from the sea to the mountains” has finally been completed, didactically and regarding the doctoral research requirements. In the following, I will discuss our research with empirical work in the field of outdoor education and I will put special focus on the methodologies applied – i.e. to the epistemological and metaphysical underpinnings of the different research paradigms. I will thus show that the historical sequence of my research projects is not arbitrary. Those papers constitute my theoretical and empirical approach to educational research “in the mirror of nature.”

1.2. Facts and values: educational research in the mirror of nature

“Contemporary educational research has been experiencing an explosion of new methodologies and approaches to inquiry. Many of these approaches have drawn from philosophical or theoretical positions that underlie their determinations of research methods, aims, and criteria of validity. Yet the substance of these philosophical or theoretical assumptions is not always made clear to the reader, and so it is difficult for one to judge those assumptions for oneself” (Biesta & Burbules, 2003).

This uneasiness about empirical methodologies in education reflects the underlying motivation for this dissertation and it applies both, to the stereotypical distinction of so-called “quantitative” and so-called “qualitative” approaches, as well as to the different methodologies within each paradigm. In the field of outdoor education, which is a relatively young academic discipline and where research is conducted from various academic fields, this Babylonian methodological confusion is even more virulent, when outdoor education is seen as both, a “research topic” and a “universal value” (Nicol, 2002a, 2002b, 2003). This distinction of “fact” and “value” in educational research, and especially in outdoor educational research, causes a vivid controversy in the educational field. It is driven by “the idea that professional practices such as education should be based upon or at least be informed by evidence“. This „continues to capture the imagination of many politicians, policy makers, practitioners and researchers“. However, as Biesta further argues, there is “a growing body of work that has raised fundamental questions about the feasibility of the idea of evidence-based or evidence- informed practice” (Biesta, 2010, p. 491).

Biesta argues against this form of “evidence-based practise”, and applies a pragmatist point of view, raising some critical epistemological, ontological, and praxeological concerns. “Evidence-based” or rather “evidence-oriented” inquiries represent the predominant strategy in educational research and it is exactly this type of research that is often in the headlines of newspapers and governs policy makers.¹

¹ This form of empirical-analytical educational science stands in the tradition of Wolfgang Brezinka’s fundamental distinction of “Erziehungswissenschaft” (educational science) as a descriptive, causalistic, and prognostic discipline, against a “philosophy of education” which is normative, and “practical pedagogics” with action-oriented theories of education (Brezinka, 1971).

“The OECD’s yearly reports (Education at a Glance) are eagerly awaited by the media and politicians, as are the publications of international comparative studies such as the Progress in Reading Literacy Study (PIRLS), the Programme for International Student Assessment (PISA) or the Trends in Mathematics and Science Study (TIMSS). Politicians want to find out about the strengths and weaknesses of their own educational system while also obtaining suggestions from international benchmarks to make policy decisions.

All this, together with an increasing tendency to consider empirical evidence when making decisions on educational policy, is based on developments in the social sciences in recent decades” (Prenzel, 2009, p. 30).

However, Biesta is „generally worried about the expectations policy makers hold about what *evidence* can and should do in relation to professional practices such as education“ and he thus wants to „provide educators with insights and arguments that can help them to resist unwarranted expectations about the role of evidence in their practices and even more so of unwarranted interventions in their practices“ (Biesta, 2010, p. 493, cursive added).

Biesta identifies epistemological problems with “evidence-based” research, i.e. that it recurs to a “representationalist theory of knowledge”, which cannot be warranted in social sciences, which he calls a “knowledge deficit” (p. 495). Secondly, he raises some ontological concerns, mainly that “evidence” is a static concept that cannot cover the complexity of social realities, and ascertains an “efficacy deficit” in “evidence-based” research (p. 497). Last but not least, he diagnoses an „application deficit“: Evidence-based research

“misses important aspects of what makes the application of such knowledge possible (particularly the work that is needed to transform the outside world so that knowledge becomes applicable) and perhaps even serves as an ideology that makes the incorporation of practices into particular networks invisible” (p. 499f)

Biesta argues that this marks, in short, a complexity reduction in “evidence-based” research and raises, despite the allegedly “neutral” focus on facts, questions of “normativity, power and values” (p. 500), when only a very narrow range of facts governs the whole field and overpowers other “realities”. This “can particularly be seen in the on-going struggles to create opportunities for ‘alternative’ medicine (and the very phrase ‘alternative’ already shows the power of what is considered to be ‘normal’)” (ibid.). In the educational field, such evidence-based practices strengthen the political significance of education by joining into the choir of effective and mainstream politico-scientific language, but it narrows the scope of educational research and practices to a small field of “controllable variables” and does not remotely capture the complexity of pedagogics.

In outdoor education, which is a much younger academic discipline, we can sense a call for more “evidence-based research” ever since the first journals started to publish exclusively about outdoor and experiential educational topics (Fengler, 2012; Neill, 2002/2008, 2006). And as early as 1984, Alan Warner found it somewhat “paradoxical that an education movement, which places so much emphasis on learn-

ing as a process focuses its research efforts on documenting products” (Warner, 1984, p. 41). I certainly join the choir of those demanding academically sharp and vigorous empirical research in the field of outdoor education. However, the very concept of “evidence” seems to be a problematic alternative in educational research, and the dualism of “qualitative” and “quantitative” research paradigms a little bit too simplistic.

In the following, I will critically examine different approaches and methodologies in outdoor education research with respect to their ontological warranty and the epistemological value of their findings. Hereby, the logical function of abduction in qualitative reasoning will be of interest in juxtaposition to deductive and inductive analysis. I will follow an interpretation of Charles Sanders Peirce’s “abductive inference” (Peirce, 1955), proposed by Lorenzo Magnani (Magnani, 2001). He distinguishes two kinds of abduction, theoretical and manipulative. Whereas theoretical abduction results in “creativity”, a rather abstract concept, manipulative abduction leads to finding methods of constructivity at the experimental stage. “Manipulative abduction happens when we are thinking *through* doing and not only, in a pragmatic sense, about doing” (Magnani, 2001, p. 53). The idea of manipulative abduction goes beyond the scope of “evidence-based” research where the empirical results form new scientific laws, as “the nature’s answers to the investigator’s question” (ibid.). Manipulative abduction refers rather to “an extra-theoretical behaviour that aims at creating communicable accounts of new experiences to integrate them into previously existing systems of experimental and linguistic (theoretical) practices” (ibid.).

Complementary to Biesta, who wants

“to make clear where pragmatism is situated in the history of Western philosophy, what its views about knowledge, reality, and human action are, and how these views might be relevant for our understanding of and approach to educational research today, both from the point of view of those who conduct educational research and from the point of view of those who use, or are affected by, its outcomes (Biesta & Burbules, 2003, p. 2f)”,

I will put forward an argument for mixed-methods approaches that allows using much of the language of “evidence-oriented”, “quantitative” empirical research without falling into the trap of metaphysical realism. This approach is much inspired by the vivid philosophical dispute between German transcendental and American analytical pragmatists, authors such as Jürgen Habermas, Karl-Otto Apel, and Hans Lenk on the one side, and Richard Rorty, Donald Davidson, and Thomas Nagel on the other. I will argue that transitions from one to another methodological paradigm are permissible, as they open up the extent of empirical research, and I will thus offer an epistemological justification of so-called and stereotypically postulated “mixed-method-approaches”.

In this experiential operationalization of pragmatist language philosophy, Rorty’s “mirror of nature” gets thus a slightly different connotation. Yet, the very concept of “nature” must be critically re-thought and freed from homonymous use:

- we use “nature” as an abstract, epistemologically realistic reference-foil for science – nowadays also in education – representing “infallible natural laws” (“evidence”);
- then there is “nature” as an object for research – in the classical “natural sciences”, where “natural” phenomena are observed, analysed, and modelled, mostly under the epistemic paradigm of “natural realism”;
- and there is the experiential field of outdoor education, where “nature” plays a contextual role and whose meaning is yet to be made explicit.

1.3. A brief history of outdoor education research

Where are we to begin with a history of outdoor education research when we do not even know what outdoor education is all about?

As Andrew Brookes laconically states in the early days of outdoor education research conferences, the meaning of outdoor education is “relative to time and place” (Brookes, 1991). Yes, indeed!

Accordingly, James Neill, one of the leading scientists in outdoor education, warns to “suggest a universal definition” of outdoor education and offers instead “terms which are commonly used and which are similar to outdoor education” (Neill, 2004) to outline the scope of its semantic function: Adventure education, adventure programming, outdoor learning, outdoor recreation, wilderness experience, adventure therapy, camping, therapeutic recreation, environmental education, and outdoor recreation.

In addition, he offers a table of 45 more and associates “major” and “minor” key words, and tabulates the terms. This chart resembles the complexity of the outdoor education concept:

- 38% of the terms refer to education or learning
- 27% of the terms refer to outdoor(s)
- 22% of the terms refer to environment, nature, or wilderness
- 18% of the terms refer to adventure
- 76% of the terms refer to other, minor keywords, the most common of which were experiential (7%), recreation (7%), camping (5%), challenge (5%), survival (5%), therapy (5%), and tourism (5%) (Neill, 2004).

Simon Priest and Michael Gass, authors of the bestselling book “Effective Leadership in Adventure Programming”, chose another strategy to “define” the field of “adventure programming” and describe the professional field instead. They state that this is “quite different from other professional vocations” and that one must work “in challenging learning environments with a variety of mediums, clients, and professional practices” (Priest & Gass, 2005 (1997), p. xi). They disseminate four major professional fields:

- Recreation programmes, that are designed to have fun, learn new activities, or become reenergised through adventure experience.
- Education Programmes, that are “aimed at understanding concepts, enriching the knowledge of old concepts, or generating an awareness of previously unknown needs through adventure”.
- Development/corporate training, that are designed to improve functional behaviours and to train “people to behave in new and different ways through adventure”.
- Therapeutic adventure programmes, which are “aimed at changing dysfunctional behaviour patterns, using adventure experiences as habilitation and rehabilitation (p. 23).

A meta-analysis of studies trying to identify professional qualities of effective outdoor-leaders performed by Simon Priest (Priest, 1987) identifies 12 core competencies that are required to work in one or more of the above mentioned professional fields as follows:

decision-making skills	effective communication
environmental skills	organizational skills
experience-based judgment	problem solving skills
instructional skills	facilitation skills
professional ethics	flexible leadership style
technical skills	safety skills

Table 1. The core competencies of an effective outdoor-leader (Priest, 1987).

Are we thus to start writing our history of outdoor education research with reference to any of the above mentioned terms, both from the associated semantic fields or from the practical context – looking for surveys on the effects on health and wellbeing of 18th century Alpine tourism or summer-camps of the *Wandervogel*-movement in the early 19-hundreds respectively studying into professional development psychology?

That might be worth a dissertation or two, but would certainly be beyond the scope of this introduction, since we would have to clarify the influence of those phenomena to what is classically associated with outdoor education, i.e. Outward Bound-style² programming. Outward Bound was the first institution in history that delivered formal out-of-doors pedagogical programmes involving challenging outdoor situations to give students the possibility to gain confidence, to redefine their own perceptions of their personal possibilities, to demonstrate compassion, and to develop a spirit of

² In Germany, for example, “Outward Bound” was used synonymously to “Erlebnispädagogik” (experiential learning) until the mid-1980ies for the lack of other institutions offering that kind of programmes.

camaraderie with their peers (cf. Richards, 2004). Thus, Neill cannot without a smile between the lines that

“the first research study, it could be said, was the first Outward Bound program. Although the statistics are not reported anywhere, the Outward Bound movement continues to claim that the survival rates of the young merchant seaman were significantly improved, and that thus the program was a success. Whether such data ever existed is doubtful, but it made a good story and helped give justification for the continuance of Outward Bound programs“ (Neill, 2006).

Neill refers to Kurt Hahn who delivered the first “Outward Bound” programme in 1941, after the joint effort with Sir Lawrence Holt, a British shipping baron, in order to teach young British sailors vital surviving skills necessary during World War II. But one could as well connote the birth of today’s Outward Bound “somewhat humbly with the opening of Gordonstoun school in Scotland in the 1930’s with only two students” (Richards, 2004). There, Hahn developed a practical curriculum which included basic elements from track-and-field (running, jumping and throwing, components of the German Sports Badge), as well as learning to step out of ordinary life through an expedition, and engaging into a project, in addition to lessons in the classroom, transferring his experiences from the Salem-school in southern Germany to Scotland (cf. Dettweiler & Kugelmann, 2010).

In the USA, outdoor education was known better in form of “therapeutic camping” or “adventure therapy” as early as 1901 with the “tent therapy” programme on the hospital grounds of the Manhattan State Hospital East to isolate TB patients from other patients. Surprisingly good experiences of patients in the fresh air lead to the development of further programmes, including milestones such as the establishment of recreational summer camps (for example Camp Ahmek, located on the mainland shore of Canoe Lake in Algonquin Park, Ontario, Canada, in 1921) or the therapeutic camping movement originated by Campbell Loughmiller in 1946 with the support of the Salesmanship Club of Dallas, before Outward Bound set foot on the American continent in the early 1950’s, too.

Betty van der Smissen (1982 (1972)) produced very helpful bibliographies of camping research and summarizes that

“[there were more than 130 master’s theses and doctoral dissertations focusing on outdoor education in this era [1950-1960 in the USA]. Almost all of these fall within the concern of operations, and may be generally classified into four types: administration, general surveys, proposed programs, and curriculum articulation. Not quite one-half of the studies are within the administration field, and include manuals, organizational structure, personnel, et al. The other studies are fairly equally divided among the other three types.

In an effort to know what was going on both nation-wide and in specific state, at least fifteen of the studies, mostly in the latter 50s, were surveys. Much of the research was very specific to one program in all of the other three types. Studies of proposed programs outlined the development and structure of a program for a given school. Many of the administration studies also were specific to operational concerns of a given school. Because these studies were so operationally-specific, they had little impact on educational research in general” (p. 117).

Finally, in the mid-1960's-1970's, Francis J. Kelly and Daniel J. Baer laid ground to research interests and designs of today's outdoor education scene, when they studied into the effect of Outward Bound programmes with delinquent youth, reporting positive, long-term benefits (Kelly, 1974; Kelly & Baer, 1968).

It is obvious, that research in such a complex field appears very diverse, and that it evolves in many academic disciplines, sometimes also in cross- or interdisciplinary research projects applying a vast variety of research methodologies (cf. Neill, 2002/2008).

However, research in outdoor education today is concentrating predominantly on effect sizes that can be measured (Fengler, 2012; Neill, 2002/2008, 2006). Peter Allison and Eva Pomeroy postulate therefore "the need for a different research paradigm" (Allison & Pomeroy, 2000, p. 96) and present a constructivist epistemology along the lines of "understanding" and propose "approaches such as ethnography, case studies, biographies and phenomenology" (p. 97). The latter methods of data collection are usually subsumed under "qualitative research methodology". So basically, according to Allison and Pomeroy, two general epistemic paradigms can be found in recent outdoor education research.

In the following, I will critically discuss those two research paradigms with respect to their epistemic values, their basic empirical operations, and their permissible statistics, and apply the analysis to one most prominent example of German outdoor education research (Fengler, 2007), as well as to my own two empirical papers. The very thorough and rigorous discussion of Fengler's dissertation reveals the main problematic implicit in so-called "evidence-based" empirical outdoor educational research. My own two empirical papers will juxtapose a different methodological and epistemological approach to the concept of evidence.

For the sub-categorization within the paradigms, I will basically follow Alf Odden (Odden, 2001) who suggested a useful classification of epistemologies in outdoor education research, and I will develop criteria of reference along the epistemologies of the applied methodologies. In his analysis of Norwegian doctoral theses in outdoor education, Odden (p. 7) differentiates between

- a) *research topic, research goal, and research problem*, looking specifically into the research project's central theme, and knowledge interests
- b) *status of the research and its perspective*, placing the project in relation to the dimension of applied research - basic research, and determining research status in relation to the concepts of exploratory, hypothesis testing and the evolving theory
- c) *theoretical basis and model use* - providing an overview of the theories and models that are most central to the thesis
- d) *methodology that has been applied* - providing an overview of the research methods used in data collection and data processing, and placing it in relation to the quantitative/ qualitative dimension
- e) *epistemological considerations* - in relation to the epistemological dimensions realism/constructivism and individualistic/collective. These two dimensions

are selected on the basis of social science textbooks in science and the current scientific discussion in sociological journals and relies particularly on the theoretical considerations that can be found in Mjøset (2000).

2. Outdoor Education Research: Merging the Two Epistemological Paradigms into Mixed-Method Approaches

2.1. The so-called “quantitative” approach: measuring “facts”

2.1.1. Measurement and scales

Within the “quantitative” paradigm,

„any subject is to find principles of numerical reckoning and methods for practicably measuring some quality connected with it. I often say that when you can measure what you are speaking about and express it in numbers you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the stage of science, whatever the matter may be” (Kelvin, 1891, p. 80f, cited from Saint-Mont, 2011, p. 6).

The quantitative approach works with the guiding intuition that information from the empirical field can be (best) expressed in mathematical terms. The simplest definition of “measurement” is “the assignment of numerals to objects or events according to rules” (Stevens, 1946, p. 677, referring to Ferguson et al., 1940, p. 340). However, in his “Notes on social measurement”, Otis Duncan (1984) argues that this definition of measurement is incomplete “in the same way that ‘playing the piano is striking the keys of the instrument according to some pattern’ is incomplete” (p. 126). With reference to Morris Cohen’s and Ernest Nagel’s “Introduction to logic and scientific method” (M. R. Cohen & Nagel, 1934, p. 294), he states that “measurement is not only the assignment of numerals, etc. It is also the assignment of numerals in such a way as to correspond to *different degrees of a quality* ... or property of some object or event” (Duncan, 1984, p. 126, cursive in the original).

But how does empirical information, which is basically “human sensation” (Stevens, 1946, p. 677) translate into numbers at all? A committee of the British Association for the Advancement of Science seriously treated this question in the 1930ies and they came up with a report considering the possibility of “quantitative estimates of sensory events”. In fact, the committee acknowledged that the “real issue is the meaning of measurement” which is “a semantic issue”, but postponed an “orderly discussion” in favour of presenting a “scales of measurement ... that fall into definite classes”. The developed “classification of scales of measurements” (ibid) are still up-to-date and applied in today’s statistics, even if there had been added one or the other scale over the years (Saint-Mont, 2011, cf. chapter 2). The problem of the semantic significance of measurements, however, remains a problem in today’s social and pedagogical science and will be addressed in the discussion of the so-called “qualitative” research in outdoor education.

Notwithstanding, Stevens (1946) proposed that measurements can be classified into four different types of scales. These are shown in the table below as: nominal, ordinal, interval, and ratio. In his “philosophy of statistics”, Uwe Saint-Mont (2011, p.

28) added one scale, the “absolute” scale, referring to Duncan’s (1984) critique of Steven’s where he argues that Steven’s “list of scale types is seriously incomplete, although we can draw on Steven’s own text, in part, to enlarge it” (p. 126). The substantial amendment Duncan proposes is to add a scale for the most simple statistical operation, “counting” (p. 151).

For each scale, Stevens defines the basic empirical operations, its mathematical group structure, and the permissible statistical operations are defined.

Scale	Basic Empirical Operations	Permissible Transformation	Permissible Statistics (invariantive)
Nominal “N”	Determination of equality	Injective function (one-to-one) $f(x) = f(x') \Rightarrow x = x'$	Number of case Mode Contingency correlation
Ordinal “O”	Determination of greater or less	Monotone function $x < x' \Rightarrow g(x) < g(x')$	Median Percentiles
Interval “I”	Determination of equality of intervals or differences	Affine linear function $h = ax + b$ with $a \neq 0$	Mean Standard deviation Rank-order correlation Product-moment correlation
Ratio “R”	Determination of equality of ratios	Linear function $h(c) = ax$ with $a \neq 0$	Coefficient of variation
Absolute “A”	Determination of equality of objects	Identity function $h(x) = x$	Number of case Counting

Table 2. The scales of measurement (Stevens, 1946, p. 678), supplemented with Saint-Mont (2011, p. 28), omitting the reference to mathematical group structure and providing information on the “permissible functions” instead.

For our purpose, it seems sufficient to concentrate on the “basic empirical operations”, “permissible transformations” and the “permissible statistics” – as those refer to the application in the research design, and omit Stevens’ reference to mathematical group theory, the more so as Stevens himself omitted reference to “group structure” in a presentation in 1975, and “the only use made of that idea was to stipulate ‘permissible transformations’ of scale values for each of the four types” (Duncan, 1984, p. 125).

The statistical operations attributed to each scale are cumulative: for example, “an interval scale can be erected only provided we have an operation for determining equality of intervals, for determining greater or less, and for determining equality (not greater and not less)” (Stevens, 1946, p. 678). Mathematically, the scales in the lower lines bear more information about the “real world” than those mentioned above; so we have a hierarchical structure of scales. Most interestingly, the nominal (N) and ordinal (O) scales allow transporting “only” qualitative information in a measurement and the broken line in Table 2 between O- and I-scales marks, in this nomenclature, the “quantum leap” from qualitative to quantitative research, implying that quantitative research is much fitter to project the complexity of the “real world” than simple “qualitative” data.

In common statistical surveys, N-scales are for example used to code gender or other unambiguous information. Gender is, in most cases, coded in a binary function where females are attributed an arbitrary number x , whereas males are identified with another number $x' \neq x$. Permissible transformations f need then to sustain this distinction, i.e. $f(x)$ must not equal $f(x')$.

O-scales contain a little bit more information. In O-scales, objects or a row of classes of objects are tabulated. It is therefore possible to unambiguously decide which object is empirically (in what sense ever) bigger or smaller than another. In mathematical nomenclature, the symbol for “empirically bigger than” is $>$, respectively $<$ for “empirically smaller than”. Thus, a monotone function g for an object a being empirically smaller than object b transforms into numerical ascriptions x and x' where $x = s(a) < x' = s(b)$, with s being the measured value (measurand) and $s \in N$. This is equivalent to $g(x) < g(x')$. Reversely, monotone functions allow to deduce $g(x) < g(x') \Rightarrow x < x'$, for $x' < x \Rightarrow g(x') < g(x)$, which is contrary to $g(x) < g(x')$ and thus unambiguously defined.

“A classic example of an ordinal scale is the scale of hardness of minerals. Other instances are found among scales of intelligence, personality traits, grade or quality of leather, etc.” (Stevens, 1946, p. 679).

In outdoor education research, we find often examples of O-scales in socio-psychological inventories when a researcher asks people to rank their preference for types of outdoor activities, with 1, for example, as the most preferred, and 4 as the least preferred (resulting in, perhaps, 1-hiking, 2-skiing, 3-rock-climbing, 4-fishing). Ordinal scales tell nothing about the distance between units of the scale (for example, although skiing may be preferred to rock-climbing, no information is available about the extent of that preference) and supply information only about the order of preference.

I-scales have equal distances between scale units and permit statements to be made about those units as compared to other units (that is, one unit may be a certain number of units higher or lower than another). Strictly speaking, I-scales allow formulating quantitative predications. But obviously, the definition of the “zero” is historically contingent. “The zero point on an interval scale is a matter of convention or convenience, as is shown by the fact that the scale form remains invariant when a constant is added” (Stevens, 1946, p. 679). Stevens quotes an example of scales of temperature, Centigrade and Fahrenheit: It is exactly defined what it means that the temperature dropped from 25° C to 15° C over night (exactly 10 degrees on the Centigrade scale), and this can be easily converted to the Fahrenheit scale. In the USA, the weather forecaster would warn: “The temperature will drop from 77° F by 18 degrees to as low as 59° F tomorrow. So be prepared and dress with an extra sweater!”

“Equal intervals of temperature are scaled off by noting equal volumes of expansion; an arbitrary zero is agreed upon for each scale; and a numerical value on one of the scales is transformed into a value on the other by means of an equation of the form $x' = ax + b$ ” (ibid).

But I-scales do not allow conclusions that one unit is a particular multiple of another because on interval scales there is no absolute zero. That is, the scale does not allow for the complete absence of the phenomenon being measured. For example, if you refer to the I-scale used on a thermometer, you can say that 40 degrees is 2 degrees higher than 38 degrees, but you cannot accurately say that 40 degrees is twice as hot as 20 degrees because there is never a situation of no heat at all. The zero on a thermometer doesn't indicate a complete lack of heat, only one more unit on the scale, which continues downward. Interval scales, then, permit a statement of “more than” or “less than” but not of “how many times more.”

In mathematical nomenclature, we can write for four measured temperatures, $x, y, u, v \in W$, respectively two measured temperature intervals $d(x, y)$, $d(u, v)$:

$$\begin{aligned} d(x, y) = d(u, v) &\Leftrightarrow s(x) - s(y) = s(u) - s(v) \\ &\Leftrightarrow a(s(x) - s(y)) + 2b = a(s(u) - s(v)) + 2b \\ &\Leftrightarrow (as(x) + b) - (as(y) + b) = (as(u) + b) - (as(v) + b) \\ &\Leftrightarrow s^*(x) - s^*(y) = s^*(u) - s^*(v) \end{aligned}$$

In I-scales, almost all the usual statistical operations are permissible, such as defining mean, standard deviation, rank-order correlation, or product-moment correlation, and allow only arithmetic operations of addition or subtraction.

In outdoor education research, we find I-scales used predominantly in simple evaluations or more elaborately, in personality inventories where certain personality traits are measured indirectly by asking the proband a set of related questions. Those are

designed in way that they can be answered by marking one definite answer on a scale.

In our survey “Investigating the motivational behaviour of pupils during outdoor science teaching within self-determination theory” (Dettweiler, Ünlü, Lauterbach, Becker, et al., 2015), we used the German adoption of a standard inventory to examine “reasons for acting in two domains to measure the grade of motivation for one” (Müller et al., 2007; Ryan & Connell, 1989). This scale works with the theoretical assumption that there exists a continuum of self-determination. There are four different types of behavioural regulation, defined in terms of the degree to which the regulation of an extrinsically motivated activity has been internalized and integrated. They are a) external regulation, b) introjected regulation, c) identified regulation, and d) intrinsic regulation, in order from the least (a) to the most (d) fully internalized. Then the questionnaire provides several possible reasons that have been preselected to represent those different styles of regulation or motivation. The respondent then checks each of the 17 propositions on a pentavalent I-scale ranging from full agreement to complete refusal.

Similar approaches using I-scales on more personality-related inquiries can be found in recent outdoor education research (cf. Duerden, Widmer, Taniguchi, & McCoy, 2009; Gookin, 2011; Passarelli, Hall, & Anderson, 2010; Sproule et al., 2013).

R-scales, finally, allow measuring periods (for example of time) “and one period may be correctly defined as double another” (Stevens, 1946, p. 679). It is obvious, that inventories trying to measure behaviour or personality traits cannot refer to an absolute or true zero point, for this would mean that the absence of *any* behaviour or of *any* personality trait was conceivable, which is, of course, not the case. It does not make any sense trying to understand the notion of “zero intelligence”. With the requirement of an absolute or “true” zero point, R-scales are most commonly used in physics where there exist operations for determining all four of the following relations: equality, rank-order, quality of intervals, and equality of ratios. As the name suggests, R-scales allow arithmetic operations such as multiplication and division. In mathematical nomenclature, one could write:

$$\frac{s(x)}{s(y)} = \frac{s(v)}{(w)} s \Leftrightarrow \frac{a \cdot s(x)}{a \cdot s(y)} = \frac{a \cdot s(v)}{a \cdot s(w)} \Leftrightarrow \frac{s^*(x)}{s^*(y)} = \frac{s^*(v)}{s^*(w)}$$

With R-scales, all types of statistical measures are permissible.

As Stevens notes, “foremost among the ratio scales is the scale of number itself - cardinal number- the scale we use when we count such things as eggs, pennies, and apples” (p. 680). He argues that this scale of the numerosity of aggregates is so basic and so common that it is ordinarily not mentioned in discussions of measurement. Duncan (1984) offers a thorough debate on the omission of cardinal numbers on the scale of measurement, their sub-grouping under R-scales, and argues to appreciate cardinal numbers with an own, absolute scale, the A-scale. In outdoor education research, A-scales are for example used to measure the dissemination of a certain phenomenon, for example the dissemination of “udeskole” in Danish schools (cf.

Bentsen, Søndergaard Jensen, Mygind, & Barfoed Randrup, 2010) or incident reporting (cf. Deutscher Alpenverein, 2011).

2.1.2. Ontological presumptions of the stochastic world of facts

Having clarified the different qualities of measurement, it yet remains to say something about the measured object's relation to the world. Since the European Enlightenment, roughly since the 16th century, Western civilization assumed that "science" could completely understand the universe and its systematic nature.

"The high-water mark of this optimistic vision came in the late nineteenth century, especially in Bismarckian Germany and Victorian England, and two of its most eloquent exemplars were Gottlob Frege, a German mathematician and philosopher, and Bertrand Russell, a British logician and philosopher" (Searle, 1999, p. 2)

Frege and Russell are in many respects anchor points of modern science. With their mathematical work, they set ground for the above elaborated scale theories, and with their logical/philosophical contributions, they outlined the contrast foil of Wittgenstein's philosophy of language, which initiated the linguistic turn in social science.

According to Searle (1999), modern science recurs to "default positions" that are not questioned and "any departure from them requires a conscious effort and a convincing argument" (p. 9). Here are (some of) the default positions implicit in standard empirical research:

- "There is a real world that exists independently of us, independently of our experiences, our thoughts, our language.
- We have direct perceptual access to that world through our senses, especially touch and vision.
- Words in our language, words like *rabbit* or *tree*, typically have reasonable clear meanings. Because of their meanings, they can be used to refer to and talk about real objects in the world.
- Our statements are typically true or false depending on whether they correspond to how things are, that is, to the facts in the world.
- Causation is a real relation among objects and events in the world, a relation whereby one phenomenon, the cause, causes another, the effect" (p. 10, cursive in the original).

Only with those default positions, empirical research, as we know it today, is possible, and most empirical researchers (including Fengler, cf. 2.2) take those default positions for granted without further reflection on their epistemological or metaphysical presumptions. As we have seen above in the example of the complex I-scale, empirical research is conducted with the aim to test hypotheses that are derived from commonly accepted theories, whose concepts are operationalized in measuring instruments as so-called variables, on their predictive power on the real world. Hereby, their reliability (for example by means of the Cronbach's Alpha test (Cronbach, 1951; Cronbach & Shavelson, 2004)) and validity must be tested in statistical spot tests that need to be independent from the spot tests of the hypotheses-test in order to

avoid circularity. Occurring positive (confirming) or negative (confutative) mathematical results decide about the explanatory power of a hypothesis (cf. Oevermann, 1996, p. 21f). Research under this paradigm aims at the examination of the “stochastically³ structured world that is in principle perceptible through the sensory channels” (p. 6, translated from German by U.D.), the world defined and described in Searle’s default positions.

According to Vollmer, one of the main characters in the field of such positivist epistemology, there are three components that influence subjective perceptions:

- the projected object (i.e. the real world)
- the projection itself
- the structure of the screen/the cognitive apparatus (cf. Vollmer, 1994, p. 121).

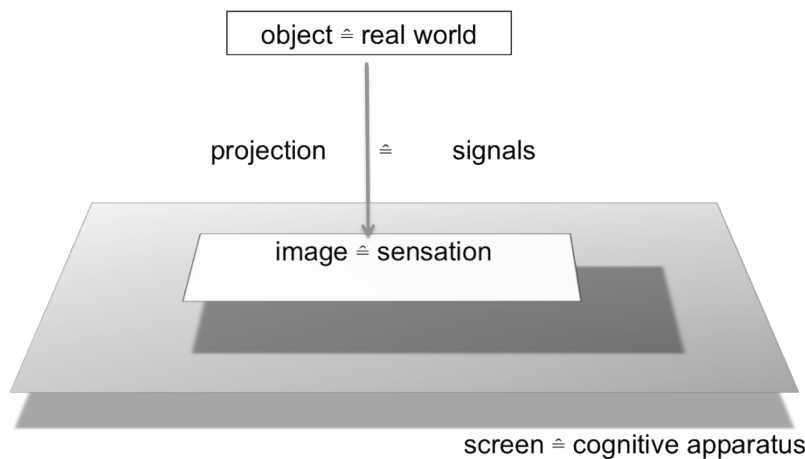


Figure 1. Projective Epistemology according to Vollmer (1994, p. 123).

Once those three components are known, the image can be determined, or rather constructed. Vollmer notes that the image will not be a 1:1 copy of the original, and speaks rather of “partial isomorphism”. The naïve realist, however, could easily drop the notion of “partial” and assume a complete projection.

If only the image is known, one can try to explain a) the original object, b) the projection mechanism, and c) the screen by applying hypotheses including the operator (↓) in **Figure 1**. There had been great philosophical debate on the projection of objects

to the mind and the “reality” of sense data. J. L. Austin, for example, put forward the standard arguments against this naïve “external realism” in his ground-breaking book “Sense and Sensibilia” (Austin & Warnock, 1962) refuting A. J. Ayer’s “Founda-

³ Ulrich Oevermann divides empirical research in a) attempts to describe the “stochastic“ world of concrete things and facts and b) attempts to reconstruct a “world of abstract, constructed meaning” (p. 234). However, the word “stochastic”, from classical Greek “stochazein”, means “to shoot with a bow at a target [...] that is to scatter events in a partially random manner, some of which achieve a preferred outcome” (Oevermann, 1996, p. 20, footnote 3; translated from German by U.D.) and it would thus be more appropriate to be used to describe a “world of possibilities”.

tions of Empirical Knowledge” (Ayer, 1961). We can well leave this debate to the historians of philosophy and need not decide for one or the other side now.

However, in the following, I will paradigmatically examine Janne Fengler’s evaluative study measuring the effects on self-concept of a series of outdoor education courses (Fengler, 2007), which is up to-date the most substantive and most acknowledged study in outdoor education in Germany. The analysis will point at some epistemological challenges of quantitative data-processing referring to the theoretical and statistical modelling applied in her research, and will show how the uncritical and unreflective use of epistemological and metaphysical default positions can taint a great research project.

2.2. Measuring “facts” in outdoor education programmes:

Janne Fengler’s evaluative study on self-concept

2.2.1. Research topic, research goal, and research problem

In her survey “Erlebnispädagogik und Selbstkonzept. Eine Evaluationsstudie” (2007), Janne Fengler offers one of the most profound and substantial explorations into measurable effects in outdoor education that are available by means of a pre-post-evaluative study with N=917 students. Fengler identifies three main research questions (pp. 3, 129, translated from German by U.D.):

RQ-1: Which changes of parameters of self-concept can be detected after the participation of an outdoor education programme?

RQ-2: What is the exact relation between change in self-concept and personal factors?

RQ-3: What is the exact relation between change of self-concept and situational factors?

The specified research population are adolescents of different age groups (under 13 (n=321), 13-14 (n=358), 15-18 (n=176), older than 18 years (n=62)) and attending different school-types (*Hauptschule* (n=16), *Realschule* (n=165), *Gymnasium* (n=300), *Gesamtschule* (n=186), *Berufsschule* (n=177), others (n=73)). Fengler’s main research goal is to examine the effect of “outdoor education” (i.e. a certain treatment) with respect to self-concept at three different measuring times to determine the significance of change after the “outdoor education treatment”.

As the text on the back cover of her book states, the context of this research problem is to produce “univariate and bivariate results, which conditions, both on participant and on the intervention sides, foster self-concept changes with adolescents in the outdoor education context”. This is the more important as Fengler is supposed to deliver “empirical proof” to “the wide-ranging field of alleged interdependency of outdoor education and self-concept change” (translated from German by U.D.).

2.2.2. Status of the research and its perspective

Janne Fengler's survey falls under the category "evaluative research" and is a splendid example of the testing of hypotheses. Fengler explicitly develops seven main hypotheses that correspond each to one of the three above quoted "main research questions".

Hypotheses H-1 and H-2 operationalize research question (RQ-1), which changes of parameters of self-concept can be detected after the participation of an outdoor education programme:

H-1: There is a significant increment of self-concept with reference to the total sample between the time of measurement t_1 and the second time of measurement t_2 .

H-2: There is a significant increment of self-concept with reference to the total sample between the time of measurement t_1 and the third time of measurement t_3 .

Both hypotheses are uni-directional and will use the whole significance level in the statistical operation rather than formulating bi-directional hypotheses and perform a two-tailed test. That presupposes that Fengler assumes to know which way the independent variable will affect the dependent variable, i.e. outdoor education *has a positive* effect on self-concept (which implies that it is theoretically excluded to have a *negative* effect).

Hypotheses H-3 to H-7 have been formulated after the examination of H-1 and H-2. Hereby, H-3 to H-5 concern research question RQ-2 asking for the exact relation between change in self-concept and personal factors, referring specifically to the variables gender (H-3), age (H-4), and school type (H-5). Hypotheses H-6 and H-7 concern research question RQ-3, asking for the exact relation between change of self-concept and situational factors, with H-6 referring to the variable "duration of the programme" and H-7 referring to the variable "type of programme". Again, all hypotheses H-3 to H-7 are uni-directional.

2.2.3. Theoretical basis and model use

Janne Fengler's evaluative study is based on two different complexes of theoretical models, (1) the learning models in (German) *Erlebnispädagogik* (and Fengler faces the same problem with defining the concept as described in chapter 1.3), and (2) the complex of "self-complex".

Theory complex (1) is divided in five segments.

- Historical overview of outdoor education
- Definitions

- Features of “erlebnispädagogische Interventionen⁴”
- *Erlebnispädagogik* as a tool of (self-concept)-modification
- Methodologies in evaluative research in *Erlebnispädagogik*.

Hereby, segments three and four are worth of deeper evaluation. Fengler understands *Erlebnispädagogik* predominantly in terms of a social pedagogical intervention. This seems to be a rather German approach to outdoor education but represents the German *Erlebnispädagogik*-market, both with reference to the German outdoor education industry with its major players *Erlebnistage -Gesellschaft zur Förderung der Erlebnispädagogik (GFE)*, *Outward Bound Deutschland e.V.*, and *Praxisfeld*, and with reference to the publication and convention market, which is dominated by *Ziel-Verlag* in Augsburg. The coining of the concept of *Erlebnispädagogik* as intervention rather than education took place in the mid 80-ies (Ziegenspeck, 1992). However, Fengler delivers a brief report about the discussion, if *Erlebnispädagogik* is to be seen as *sub-discipline* in the academic canon of subjects (cf. for example Düchting, 1994), or rather as a *method* (Bedacht, 1994; Heckmair & Michl, 2012 (1993); Hufenus, 1997, 2009) and is very reflective on her own position seeing *Erlebnispädagogik* as both, academic sub-discipline *and* method (cf. Fengler, 2007, p. 19) and uses the concept accordingly in her research design.

More problematic is the forth segment, conceptualizing *Erlebnispädagogik* as a tool of (self-concept)-modification. After identifying core features of *erlebnispädagogische* programmes from literature, quoting Reiners and Schmieler (1997), Carver (1996), Gilsdorf (1999), and Galuske (2003), Fengler refers to three reflexion models in *Erlebnispädagogik* which Steven Bacon introduced in 1987 as historically emerged (Bacon, 1987). Bacon’s essay, “The Evolution of the Outward Bound Process”, where he “examines the way in which the Outward Bound process has evolved in the United States with particular emphasis on how it has changed to ensure greater transfer of course learnings“ (p. 1) has been, despite it had never been properly published, most influential to theory building in outdoor education. It has been picked up by Simon Priest and Micheal Gass in their above mentioned book “Effective leadership in adventure programming” (Priest & Gass, 2005 (1997)) in its first edition in 1997, and has been introduced in Germany mainly by Heckmair and Michl (2012 (1993)), Schaad (1993), and Schödlbauer (1999, 2000) and perpetuated in countless bachelor’s, master’, and doctoral theses. Bacon’s

“typology of curriculum models is developed consisting of: 1) a first generation model—focusing on experience alone—which dominated Outward Bound programming in the 1960's and early 1970's, 2) a second generation model—emphasizing discussion, group process, and imported techniques—which is the current ruling paradigm at Outward Bound, and 3) a third generation model—stressing experiential metaphors—which may provide a direction for future curriculum evolution. The three models are contrasted, and the strengths and weaknesses of each are explored“ (Bacon, 1987, p. 1).

⁴ I quote the original German because Fengler enters into the chorus of social pedagogues in Germany who understand “outdoor education” predominantly in terms of “intervention”, rather than of “education”.

In the German discourse, those three models are called “The Mountains Speak for Themselves” (English in the original), the “Outward Bound Plus” model, and the “Metaphorical Model”, and it is assumed, that those models have historically emerged over the last 50 or so years in outdoor education practice. Bacon, however, is using his categorization only to illustrate the evolution of the Outward Bound process in the USA, and here pre-dominantly within the Colorado Outward Bound School in Denver. The only source of Bacon’s elaboration of his first model is an unpublished manuscript written by the Outward Bound instructor Thomas James in 1980, which was made accessible by James Neill on his website in 2007 (James, 1980). Bacon quotes big chunks of Thomas’ essay:

“To begin with, it seems to me that people who are saying anything equivalent to ‘Let the mountains speak for them-selves’ are also saying something more, which is that instructors can rely on the overall structure of the Outward Bound course to give their students a good experience. They can rely on a training sequence, a way of grouping students and committing them to task performance, activities like solo and the rappel, etc. ... So the point is not exactly that the mountains do the teaching. It is that the training sequence we are using is a remarkably effective way to get people to learn in the mountains... The experience happens naturally if instructors are skilled enough to take their students safely through the adventurous activities that make up Outward Bound, and when they do that, the mountains are extraordinary teachers indeed” (quoted after Bacon, 1987, p. 3).

It is obvious, that James describes a certain attitude of Outward Bound instructors known to him at a certain period of time. The essay is rather narrative in character, reporting that “several years ago” (from 1980, the time, the essay was written)

“a course director named Rustie Baillie coined the phrase, "Let the mountains speak for themselves". He was reacting against pressures in the Colorado Outward Bound School to verbalize student experience on courses and to use counselling techniques to manage the group process of patrols. Baillie was not the first to react. In interviewing staff and trustees from the early years of Outward Bound in this country, I discovered that the issue was as vehemently debated then as it is today. In fact, the debate began right after the first season in 1962 when there was a falling out about whether to instil an "intellectual element" in courses, and since then there have been plenty of historical examples of the rift. In 1964 the school director required nightly staff debriefings on the meaning of each day’s activities in the base camp at Marble, and one staff member of those years told me that instructors breathed a sigh of relief when it came time to go on the six-day alpine expedition. In 1966 the chief climbing instructor tried to introduce a written guide to counselling techniques. In 1967 or so there started to be readings and other written materials available for use in the field. By the next year there was an outline to help foster spiritual awareness” (James, 1980, p. 1).

The generalization, however, that this was the general model used in outdoor education at that time, and that the historical “Evolution of the Outward Bound *Colorado* Process” is paradigmatic, remains completely unwarranted by any historically sound methodology. And the extrapolation of assumptions on the status quo of Outward Bound instruction at the Outward Bound Colorado School to the German context is utterly absurd and flies in the face of reason. The same can be shown with the other two models, but fortunately, Fengler does not make use of this typology other than

uncritically quoting it. The typology of her treatment is much more profane and less theoretical. It relies only on (1) the course contents offered by *Praxisfeld* (2002, 2006), the company whose programmes were evaluated in the survey:

- experience of body and space
- training of senses
- trust-building
- co-operation
- reflexion/transfer

and (2) on rules of conduct, which include

- challenge by choice
- to welcome getting involved with something novel
- the stop-rule
- safety procedures (cf. Fengler, 2007, p. 137).

She further specifies the course-types according to duration and courses containing a high ropes-course-element (RC) and courses that do not (non-RC). As mentioned, there is no reference to any of the above quoted theoretical models, but it is somehow implied that by applying all of the above mentioned features in “naturally very divergent weight according to the specific tailor-cut exercises” (p. 136), typical samples of *Erlebnispädagogik* are represented and can thus be compared. So despite her 50 pages of theoretical model contemplation, the model finally applied boils down to sheer pragmatic considerations that are so general that they resemble rather a didactical black-box.

Theory complex (2) refers to self-concept and is much more stringent and evidence based. After a profound discussion of the self-concept research literature, which she divides into “basic research” and “evaluative research”, Fengler chooses to recur to Ingrid Deusinger’s *Frankfurter Selbstkonzept-Skalen* whose internal reliability coefficient Cronbach’s Alpha lies between .70 and .88 and whose standard-deference values S are excellent ($0,6 \leq S \leq 0,8$) (Deusinger, 1986). Fengler’s survey shows similar results with a Cronbach Alpha range of .50 and .83 and a standard-deference of .7 and .8 (Fengler, 2007, p. 148f). Deusinger’s *Frankfurter Selbstkonzept-Skalen* have furthermore been evaluated by Asendorpf and Aken (1993) in the course of their own longitudinal study on self-concept and been discussed with other German and international self-concept scales.

2.2.4. Methodology that has been applied

In contrast to the above quoted example, Fengler can address all her research questions with one and the same instrument, since the research questions are interdependent. The research design is construed as a longitudinal pre-post-post test in a quasi-experimental field setting. Fengler identifies her design as a “theory oriented evaluation” with a special focus on “practical aspects” in form of a “micro-evaluation” (Fengler, 2007, p. 139, translated from German by U.D.).

Fengler discusses the desirability of an independent control group but concedes that for logistical and theoretical considerations, such cannot be realized:

“A control group of the required size, without communication with the members of the test group and with a similar, but less effective treatment (‘from the treatment independent intervention’, (Brezing, 2000, p. 15) is impossible to realize with respect to the sampling procedure and the expenditure of cost and time within a PhD-project. Furthermore, the comparability of the two test groups would remain doubtful in a control group design with respect to independent variables *in toto* as well as with respect testing the differential hypotheses in the diverse partial samples” (Fengler, 2007, p. 140, translated from German by U.D.).

Fengler offers a research design that includes a number of compensation measures:

The first test sample t_1 is taken immediately before the outdoor educational treatment, normally shortly after the class has arrived at the seminar location. The second test sample t_2 is taken at the end of the treatment before the debriefing, according to the specific programme after 1.5 – 3.5 days. The third sample is taken four weeks after the treatment in order to allow significant changes in self-concept, which can be assigned with high probability to the influential factor “outdoor educational treatment”. Additionally, participants of the shorter programmes can be seen as a control group to those who participated on the longer seminars.

The inventory used, the *Frankfurter Selbstkonzeptskalen* (FSKN), differentiates between four “important realms of the self”, i.e.

- a. performance, measured in four sub-scales,
- b. self-worth, measured in one sub-scale,
- c. mood and sensitivity, measured in one sub-scale, and
- d. psycho-social realm, measured in four sub-scales (Deusinger, 1986, p. 6).

Fengler focuses on three sub-scales that are taken from the three major realms a, b, d, and combines the three in one questionnaire, whereby the original unsystematic-alternating sequential arrangement of items from different scales is accounted for (cf. Deusinger, 1986, p. 34; Fengler, 2007, p. 134f).

The items are marked on an interval scale with a choice of six levels. Hereby, the 6-ary-scale is designed in a way that checking a high number on some of the items resembles a positive self-concept value (for example items 1, 6, 7) and vice versa (2, 3, 4, 5). This procedure guards against socially desired or automatized answering. In the data processing, those items have been transliterated accordingly to associate positive self-concept features with high numbers.

In addition to the I-scales from FSKN, the following relevant structural variables have been enquired: age, gender, class, and school type.

The data collection has been undertaken by different people, mainly by outdoor instructors (t_1 and t_2) and teachers (t_3). This requires special attention with respect to procedural objectivity as Fengler concedes with reference to Bortz, Döring, and

Bortz (1995). Fengler meets this challenge by having produced a handbook for instructors and teachers in order to keep the conditions constant. With respect to sampling t_1 and t_2 , she reports an “optimal” realization of standardization, but must admit certain irregularities with t_3 : here the sampling took place in an interval of three and five weeks, due to school holidays (Fengler, 2007).

The data interpretation is carried out as statistical analysis in both, longitudinal and cross-sectional comparisons by applying diverse mathematical operations on grouped subscale-means.

Pearson's Chi-squared Test is used to assess two types of comparison: tests of goodness of fit and tests of independence was developed during the period 1891-1900. Pearson's 1900 paper on chi squared “is one of the great monuments of twentieth century statistics” and “everyone agrees that such procedures now form an essential part of statistical method” (cf. Böhm-Kasper, Schuchart, & Weishaupt, 2009, p. 131ff; Plackett, 1983, p. 70; 59).

The test of goodness of fit helps to decide whether or not an observed frequency distribution differs from a theoretical distribution. The latter assesses whether paired observations on two variables, expressed in a sufficiently formed contingency matrix, are independent of each other (in Fengler's case: if person-related variables as gender, age, and school type are independent from situational variables such as programme type and duration).

	Data	Features	Test
Gender-specific differences in self-concept		Good overview of global differences	
Age-specific differences in self-concept			
School-type-specific differences in self-concept	Grouped sub-scale medians (3 means for each of the 3 subscales at 3 samples (t_1, t_2, t_3) measured in I-scales transformed into an O-scale)	Comparability of <i>univariate</i> and bivariate data	Chi-square-Test (χ^2) (Pearson, 1900)
Bivariate correlations		Good overview of reinforcing and weakening effects Verifiability of <i>bivariate</i> correlations through forming a sufficient matrix	

Table 3. Fengler's design for the cross-sectional examination (cf. Fengler, 2007, p. 152, translated from German and modified by U.D.).

The first step is to calculate the chi-squared test statistic, χ^2 , which resembles a normalized sum of squared deviations between observed and theoretical frequencies. The second step is to determine the degrees of freedom, i.e. the number of values in the final calculation of a statistic that are free to vary, usually notated with the lower-case Greek letter ν . This is essentially the number of frequencies reduced by the number of parameters of the fitted distribution. In the third step, χ^2 is compared to the critical value of no significance from the χ^2_ν distribution, which in many cases gives a good approximation of the distribution of χ^2 .

Fengler sub-scales the 6-item I-scale into three classes: values under 4, values from 4 to 4.9, and values bigger than 4.9, transforming hereby the I-scale into O-scale values and defining thus the sub-scale medians.

The longitudinal approach addresses hypotheses 1-7. H-1 and H-2 (cf. page 24f) by comparing the sub-scale means, hypotheses 3-7 had been calculated with sub-scale mean-differences. The calculation of bivariate correlations was performed with the grouped sub-scale median-differences (as described above).

All of the four mentioned tests are standard applications in statistics software, such as SPSS, which was used by Fengler in Version 12.0 (cf. Fengler, 2007, p. 154). Operating with sub-scale means to test hypotheses H-1 and H-2, which are *global* longitudinal hypotheses, is conventional procedure.

In order to test longitudinal hypotheses that are *differential*, differences in reinforcing and weakening effects can be calculated with reference to the latitude of measured effects in the sub-samples. Here, different tests are offered by SPSS: The Wilcoxon signed-rank test is a non-parametric statistical hypothesis test used when comparing two related samples, matched samples, or repeated measurements on a single sample to assess whether their population mean ranks differ (i.e. it is a paired difference test). The Mann-Whitney U Test is used to compare differences between two independent groups when the dependent variable is either ordinal or interval but not normally distributed. The Kruskal-Wallis Test is an extension of the Mann-Whitney Test to allow the comparison of more than two independent groups. It is used when three or more sets of scores that come from different groups, shall be compared (cf. Böhm-Kasper et al., 2009, p. 100ff; cf. Bühl & Zöfel, 2005).

Finally, Pearson's Chi-square-Test is used again to calculate the bivariate correlations of reinforcing and weakening combinations of variables with respect to dependent variables also in the longitudinal design.

	Data	Features	Test
H-1 and H-2: Changes in self-concept t_1-t_2, t_1-t_3	Sub-scale means	Good overview of global differences Comparability of <i>global</i> longitudinal hypotheses	Wilcoxon-Test (Wilcoxon, 1945)
H-3: Changes in self-concept "gender"			Mann-Whitney-U-Test (Mann & Whitney, 1947)
H-4: Changes in self-concept "age"			Mann-Whitney-U-Test
H-5: Changes in self-concept "school type"	Sub-scale mean-differences, each t_1 - t_2 and t_1 - t_3	Good overview of reinforcing and weakening effects Comparability of <i>differential</i> longitudinal hypotheses	Kruskal-Wallis-H-Test (Kruskal & Wallis, 1952)
H-6: Changes in self-concept "duration of programme"			Mann-Whitney-U-Test
H-7: Changes in self-concept "type of programme"			Mann-Whitney-U-Test
Bivariate correlations	Grouped sub-scale median - differences	Good overview of reinforcing and weakening combinations of variables with respect to dependent variables Verifiability of <i>bivariate</i> correlations through forming a sufficient matrix	Chi-square-Test (χ^2) (Pearson, 1900)

Table 4. Fengler's longitudinal research design (Fengler, 2007, p. 153, translated from German and modified by U.D.)

2.2.5. Epistemological considerations

Janne Fengler's study on "Erlebnispädagogik und Selbstkonzept" does not question any of Searle's default positions. And the data-collection process posed no severe logistical problem of either (Fengler, 2007, p. 146). The questionnaires had been properly filled in by the participants "according to the instructions" at t_1 and t_2 ($867 \leq n \leq 900$) and most of the participants were also present at t_3 ($816 \leq n \leq 820$) four weeks after the course. The interval related to the three sub-scales that had been divergently filled out by participants ($N=917$)(p. 316). But both, Fengler's empirical and statistical modelling raise some epistemological concerns.

Pearl (2009, p. 202) defined a model as "an idealized representation of reality that highlights some aspects and ignores others." In the following, I will analyse Fengler's *empirical* model.

Fengler's empirical model

In Fengler's empirical approach, any changes of self-concept are *indirectly* measured by means of a questionnaire composed of items that indicate in one or the other direction, but do not directly reflect the outdoor educational setting itself. Statements as "I am afraid of being myself" (item 7, Fengler, 2007, p. 312) or "I feel contempt for myself" (item 10, p. 313) may reliably and validly hint at a person's self-concept if rated correctly, and differences in value may be calculated by means of the most elaborate statistical operations, yet, not one of the calculated features could be verified or exemplified with a single observed reference to the "real world".

This is noteworthy since Fengler abjures *direct* field-observation at all and leaves the only field-related research-act, i.e. the issuing of the questionnaires, to outdoor instructors and teachers. Those are two operations, which put the researcher into the furthest possible distance to the research field. Guba characterizes the empirical model behind that sort of rationalistic "disciplined inquiry in education", which was coined by Cronbach and Suppes (1969), as follows:

„There is a single, tangible reality fragmentable into independent variables and processes, any of which can be studied independently of the others; inquiry can converge on this reality until, finally, it can be predicted and controlled“ (Guba & Lincoln, 1982, p. 237).

Anyone being able to use statistics software and being familiar with this sort of research algorithms could have conducted this survey using psychological inventories which „appear to be more cost efficient, they have a patina of objectivity, and they produce information that can be systematically aggregated“ (Guba & Lincoln, 1982, p. 245). Objectivity is, in any such "disciplined inquiry in education", one of the main quality criteria. However,

“in general, the objectivity of empirical social research is an objectivity of the methods, not of what is investigated. From surveys of varying numbers of individuals,

statements are derived by means of statistical processing which are generalizable and independent of individual fluctuations in accordance with the laws of the theory of probability.” (Adorno, 1976b, p. 71).

Adorno uses this characterization of empirical research in sociology as a contrast foil to hint at the “societal objectivity, the embodiment of all the conditions, institutions and forces within which human beings act, or at most, they have taken them into account as accidentals” and states that “at fault here are not only those interested in commissioning research who consciously or unconsciously prevent the elucidation of such conditions“ (ibid).

But Oevermann’s account of the “stochastic world” (Oevermann, 1979b, 1996) helps us to understand the epistemological presumptions implicit in such sort of empirical research. The interpretation of the “observed” data in Fengler’s survey is simultaneously defined with the design of the data collection. You need not be an expert in outdoor education in order to read out the data on self-concept from the SPSS-tables. Neither need you be an expert on self-concept theory, as long as you don’t design your own questionnaire. It is sufficient to apply the *Frankfurter Selbstkonzeptskalen* correctly and do your math, i.e. apply SPSS correctly. Fengler, of course, shows her expertise in both fields, outdoor education and self-concept theory by elucidating the academic discourses, but that has of course no effect on the data, and little relevant effect on the data-interpretation. As we have seen in the analysis of Fengler’s discussion of the different models in outdoor education, it does not help us much to understand any of the obtained data. Its rhetorical function is merely to show competence where it is – methodologically speaking – not effectively applicable, since none of those models are relevant for the examined course-designs.

Oevermann argues that in such a “subsumption –logical” empirical approach

“the data interpretation is predetermined by the operationalization of the used concepts and the mode of data-collection and is thus performed mechanically. An independent interpretational problem begins only after the statistical testing of the hypotheses respectively after the inductive configurative calculations of factor- and cluster-analyses. Such interpretations however are external to the statistical preparation of data and relate to the latter like artefacts” (Oevermann, 1996, p. 20, footnote 3; translated from German by U.D.).

Saint-Mont offers a simple and clarifying empirical model (Saint-Mont, 2011, p. 72f) with three (partially) independent steps:

First, data have to be collected in the empirical field/structure, i.e. *measured or scaled* (↑). The results of such scaling are, in most cases, *numerical relatives*. Those are, in a second step, statistically analysed (⇒) and transformed into *numerical results*. As Saint-Mont emphasizes, those statistical operations should only be applied, if the empirical field is well represented in the numerical relatives. Otherwise, the results are nonsensical. Mathematics is only an auxiliary discipline that helps to extract information from the data and to reduce a huge junk of data into some diagnostically conclusive numbers. Finally, the numerical results are

re-interpreted and referred back (↓) to the real world and transformed into *interpretational results*.

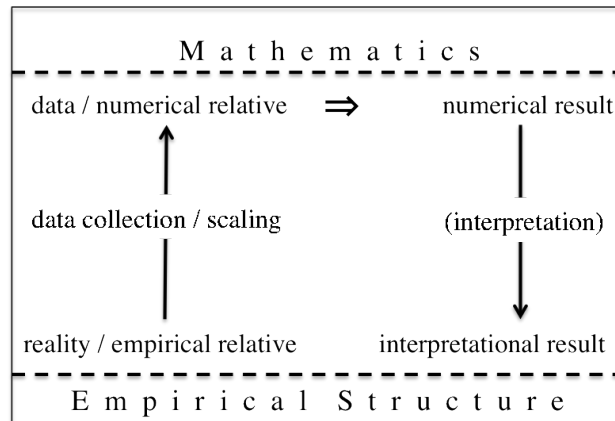


Figure 2. Empirical structure and mathematics. Model according to (Saint-Mont, 2011, p. 72).

In this empirical model, the researcher’s subjective cognitive apparatus is eliminated and replaced with the whole field of mathematics – given that the empirical structure is such that it can be fully projected into numbers and numerical relations. The advantages of such an empirical model is evident: The empirical field can be mathematically structured and fully interpreted by machines/computers (cf. Church, 1936; Turing, 1937). And given “that mathematics is the most precise realm of knowledge known to man” (Saint-Mont, 2011, pp. 13, footnote 23, translated from German by U.D.), empirical research of such sort *must* produce by default the most exact results, as it is assumed by Fengler applying the *Selbstkonzeptskalen*.

But there lies one big danger, as the famous statistician John Wilder Tukey remarks in a 1961 essay: In order to achieve good results, one needs a clear-cut problem, since “unquestioned assumptions lead to secure conclusions.” But

“[h]ow does one make a clear-cut problem? By choosing very precise and clear-cut hypotheses. And how does one solve problems completely? By wringing everything possible out of these hypotheses. These are basic mathematical techniques, essential in their place. Every clear-cut problem is artificial, separated from the real world by idealization after idealization” (Tukey, 1986 (1961), p. 153).

As stated above in Adorno’s quote, the achieved accuracy and objectivity is merely an “objectivity of method” and nothing can be said beyond the numerical values without leaving the realm of numbers. The important phrase in Tukey’s quote is that mathematical techniques are essential *in their place*. That means, they are misleading when used *out of their place*. But most often, or at least in outdoor education research, it is exactly there, where the interesting part of research begins. I want to illustrate this with one example taken again from Fengler’s study.

The results of the significance test of the differential change of self-concept with reference to programme-type, i.e. programmes including a ropes-course element (RC) and programmes without (non-RC), are surprising: in both time dimensions (t₁-

t₂; t₁-t₃), participants of non-RC-programmes achieve higher re-enforcing effects. Those results contradict the commonsensical assumption that programmes, that involve a (high-) ropes course element, effect the participants' self-esteem in a positive way. However, Fengler gives a substantial explanation on her "remarkable discovery" (Fengler, 2007, p. 230): RC- programmes concentrate more on the individual, whereas non-RC-programmes focus more on group-activities. The challenges within a group seem to have a stronger effect on self-concept than the individual challenge in a high-ropes course. In her interpretation, she refers mainly to Sarah Leberman's and Andrew Martin's 2003 paper, which present results

"from two Outward Bound studies, one for international participants in the Czech Republic and one for female offenders in New Zealand. Open-ended questions asked participants which activities took them out of their comfort zones and from which activities they had learnt the most. The experiential learning literature advocates that by moving people out of their comfort zones learning takes place. However, the findings indicate that the activities, which participants identified pushed them out of their comfort zones, may not necessarily be the activities that result in peak learning experiences" (Leberman & Martin, 2003, p. 10).

It is noteworthy that Leberman and Martin performed a "qualitative" survey and gained knowledge from asking the participants open-ended questions in a questionnaire six months after the course. The responses were

„then analyzed using the HyperRESEARCH qualitative data analysis package (Researchware, 1998), which involved the primary researcher coding the responses. These codes were then grouped and linked together into main themes (A. J. Martin, 2001). Themes related to physical, creative, group, or individual activities. It is acknowledged that other themes could have been elicited from the results. The qualitative approach did not attempt to quantify statistically the effects of the course. However, as Burnard (1991) recommended, the data was coded, categorized and 'quantified' in order to present the findings“.

One may well question the quality of the answers of Czech delinquents to researchers from New Zealand six months after the original experience, but what is important here is that the narrative data stand in *immediate* relation to the participants' experience in the field and are *not mediated* via a theoretical construct of a psychological inventory.

It is also noteworthy, that Fengler refers in the continuation of her discussion to

"findings of [her] own field work ... that non-RC-programmes display a higher density of experience by group interaction than RC-programmes, which probably leads to more frequent or more meaningful possibilities of experiencing or oneself in a new way" (Fengler, 2007, p. 232f, translated from German by U.D.).

However, we do not learn anything more specific about the quality of the RC-programmes examined by Fengler – if they can be compared to her own field-experience at all, etc. We are offered only one hint that the two types of programmes were "similar" while applying the principle of didactic reduction, i.e. designing a programme according to increasing difficulties (p. 234).

This is surely a very subjective reference and methodically insufficient approach to explain the differences in effects on self-concept in both programme types. It is, however, a step in the right direction.

Bay the way: A meta-analysis of 44 studies “that examined the impacts of participation in challenge (ropes) course activities” (Gillis & Speelman, 2008, p. 111) suggests positive impacts of such activities on the participants and contradicts Fengler’s and Lebermann & Martin’s counter-intuitive explanation. And in their meta-analysis, Gillis and Speelman (2008) conclude that “implications for further research include the importance of recording detailed program design information, selecting appropriate instrumentation, and including follow-up data” (p. 111).

Without solid information of such first hand, field-related and observed “quality”, any methodological objectivity produces only nonsensical numbers and merely speculative interpretations. The semantic significance of measuring needs to be taken into account when we want to understand something at all in empirical analyses.

„The onus of objectivity ought, therefore, to be removed from the inquirer and placed on data; it is not the inquirer's certifiability we are interested in but the confirmability of the data“ (Guba & Lincoln, 1982, p. 247).

“Neue Forschung braucht das Land!” – “new research is needed in the country” – is the title of a more recent of Janne Fengler’s articles (Fengler, 2012). Yes indeed, but not without providing data from direct and well defined field observation!

Fengler’s statistical model

After having analysed Fengler’s empirical model, it is epistemologically interesting to see how it relates to her statistical model. As Rodgers states in his 2005-address to the annual meeting of the Society of Multivariate Experimental Psychology, “the definition of a model has two important, characterizing, features:

1. A model matches the reality that it describes in some important ways.
2. A model is simpler than that reality“ (Rodgers, 2010, p. 5).

A statistical model is one that captures the above two features within one or more mathematical equations. Luce (1995, p. 2) suggested that “mathematics becomes relevant to science whenever we uncover structure in what we are studying” and that one “should not underestimate the difficulties in isolating such structure and the even more difficult task of finding good ways to describe it.” The tension embedded within this definition is obvious. The better a model matches reality, the less simple it becomes. Or, as it becomes simpler, it necessarily loses some of its match to reality. In current educational research, this tension is resolved statistically, mainly by applying null-hypothesis significance testing (NHST) in the tradition of Neyman and Pearson (1933) and Fisher (1925). The NHST-model, that is used by Fengler as well, has been generally criticized since the early 1950s and most prominently been refuted by Jacob Cohen in his essay “The Earth Is Round ($p < .05$)” from 1994 (J. Cohen, 1994). His main criticisms on the NHST-model include:

1. With a large enough sample size, all null hypotheses can be rejected.
2. Rejecting the null does not provide logical or strong support for the alternative.
3. Failing to reject the null does not provide logical or strong support for the null.
4. NHST is backwards, because it evaluates the probability of the data by means of the hypothesis, rather than the probability of the hypothesis by means of the data.
5. Statistical significance does not imply practical significance (cf. Rodgers, 2010, p. 3).

As Cohen criticizes, proponents of NHST statistics commit a fundamental logical mistake by claiming that if the null hypothesis H_0 is rejected, then the theory is established. Their standard argument is construed as follows:

“If A then B; B therefore A. But even the valid form of the syllogism (if A then B; not B therefore not A) can be misinterpreted. Meehl (1990a, 1990b) pointed out that in addition to the theory that led to the test, there are usually several auxiliary theories or assumptions and ceteris paribus clauses and that it is the logical product of these that is counterpoised against H_0 . Thus, when H_0 is rejected, it can be because of the falsity of any of the auxiliary theories about instrumentation or the nature of the psyche or of the ceteris paribus clauses, and not of the substantive theory that precipitated the research. So even when used and interpreted "properly," with a significance criterion (almost always $p < .05$) set a priori (or more frequently understood), H_0 has little to commend it in the testing of psychological theories in its usual reject- H_0 -confirm-the-theory form.” (J. Cohen, 1994, p. 999).

It is worthwhile to look at the core of this argument, since it is a widespread belief that “the level of significance, at which H_0 is rejected, say .05, is the probability that it is correct or, at the very least, that it is of low probability” (p. 998). NHST argues that if the null hypothesis is correct, then these data are highly unlikely. Now, if some of these data have occurred, the null hypothesis is highly unlikely. Cohen argues that by making the syllogism probabilistic, it becomes invalid: - Why?

“Well, consider this:

The following syllogism is sensible and also the formally correct modus tollens:

If a person is a Martian, then he is not a member of Congress.

This person is a member of Congress.

Therefore, he is not a Martian.

Sounds reasonable, no? This next syllogism is not sensible because the major premise is wrong, but the reasoning is as before and still a formally *correct modus tollens*:

If a person is an American, then he is not a member of Congress.

(WRONG!)

This person is a member of Congress.

Therefore, he is not an American.

If the major premise is made sensible by making it probabilistic, not absolute, the syllogism becomes formally incorrect and leads to a conclusion that is not sensible:

If a person is an American, then he is probably not a member of Congress. (TRUE, RIGHT?)

This person is a member of Congress.

Therefore, he is probably not an American. (Pollard & Richardson, 1987)

This is formally exactly the same as

If H_0 is true, then this result (statistical significance) would probably not occur.

This result has occurred.

Then H_0 is probably not true and therefore formally invalid.

This formulation appears at least implicitly in article after article in psychological journals and explicitly in some statistics textbooks - ‘the illusion of attaining improbability’ (J. Cohen, 1994, p. 998).

Cohen calls this an “illusion of attaining improbability” and warns the defenders of NHST to “don’t look for a magic alternative, ... some other objective mechanical ritual to replace it. It doesn’t exist” (p. 1001).

Accordingly, Rodgers concludes, that statistics needs an epistemological shift from applying “rule-based” models “mechanistically” to more creative modelling, which he sees as a “quiet revolution”:

“Engaging in science as a creative process requires thinking scientifically in creative ways. The application of NHST as a mechanical set of procedures precludes creativity. Building and evaluating statistical and mathematical models encourages creativity. To support the broad epistemological shift within the field of psychology from the first to the second perspective is the primary goal of this article, a process well underway, but one that requires further specification, organization, and attention” (Rodgers, 2010, p. 10f).

In her research design, Fengler applies exactly this sort of “pre-revolutionary NHST-approach” to a considerably huge data-set with her rather simplistic model reducing the complex pedagogical processes in outdoor education to three subscales on a questionnaire and calculating standard-deviations to an assumed normal distribution with an assumed significance level of $p \leq 0.05$ (Fengler, 2007, p. 155). In such design, the null hypothesis is rejected in favour of the specified alternative, like in a court case, if the data patterns are inconsistent enough with the distributional predictions of the null. Such model does neither capture the complexity of the empirical model – which, as we have seen in Fengler’s discussion of RC- versus non-RC-programmes, should contain also “qualitative” data directly observed from the field – nor is it able to account for causal relations – which, as we have seen, would be desirable for statistical models and data interpretation.

“The word cause is not in the vocabulary of standard probability theory. It is an embarrassing yet inescapable fact that probability theory, the official mathematical language of many empirical sciences, does not permit us to express ... causal explanations” (Pearl, 2009, p. 134).

Alternatively to her statistical model, Fengler could have used structural equation modelling (SEM) instead. Her data set was surely big enough for complex modelling, and SEM has developed a language that can project causal relations in mathematical equations. SEM is such an alternative, post-revolution approach that does also remedy the null hypotheses paradox by not any longer putting the focus on the null hypothesis but on the applied statistical model itself.

“Rather than testing null hypotheses that our data arise from chance process, we should be developing mathematical models and evaluating them statistically. ... The best model is the one that fits the data best in relation to its complexity. ... A prototypical modelling approach, the development of which has helped define the modelling revolution, is structural equation modelling (SEM)” (Rodgers, 2010, p. 5).

Structural equation modelling is a blend of *confirmatory* procedures, where the construct of the model is examined on consistency with a researcher's understanding of the nature of that construct, and of *exploratory* procedures, where the underlying structure of a (relatively large) set of variables can be tested, which shifts the focus away from the null hypothesis (Kline, 2011, p. 9ff). “In a sense, the null and alternative hypotheses still exist, but the null now has flexibility” (Rodgers, 2010, p. 5), and the evaluation of model fit and test of the hypotheses is still performed with a robust version of Pearson’s χ^2 test, for example Satorra-Bentler scaled χ^2 (Satorra, 2000; Satorra & Bentler, 2010). It is thus a statistical technique for testing and estimating causal relations using a combination of *statistical data* and *qualitative causal assumptions* (Kline, 2011; Pearl, 2009).

It would be interesting to see, how Fengler’s results, for example with respect to RC- and non-RC-programmes, would look like when examined with structural equation modelling, after this “silent methodological revolution” (Rodgers) has affected recent educational research (cf. Bauer, 2008; Bauer & Prenzel, 2012).

An alternative to the SEM-approach (which would, by the way, still be possible ex-post), would have been a mixed-methodology approach with the

“inclusion of issues and strategies surrounding methods of data collection (e.g., questionnaires, interviews, observations), methods of research (e.g., experiments, ethnography), and related philosophical issues (e.g., ontology, epistemology, axiology)” (Johnson, Onwuegbuzie, & Turner, 2007, p. 118).

Such an approach would aim at understanding the meaning of data and to comprehend social structures behind and beyond numbers, which is paradigmatically claimed by so-called “qualitative” researchers and which will be again critically examined.

2.3. The so-called “qualitative” approach: aiming at understanding meaning

2.3.1. The significance of scales

According to Steven’s theory of measurement (Table 2), so-called “qualitative research” refers to nominal and ordinal scales (cf. Stevens, 1946). The nominal scale is, according to Stevens,

“a primitive form, and quite naturally there are many who will urge that it is absurd to attribute to this process of assigning numerals the dignity implied by the term measurement. The ordinal scale arises from the operation of rank- ordering. Since any 'order-preserving' transformation will leave the scale form invariant, this scale has the structure of what may be called the isotonic or order- preserving group” (Stevens, 1946, p. 679).

It is obvious, that this pejorative classification of “qualitative scales” is due to an epistemological bias. In Steven’s positivistic view, “the assignment of numerals to things so as to represent facts and conventions about them” is the only way to obtain scientific knowledge. And knowledge that is so simple that it can only be scaled in nominal or ordinal form is, in its literary sense, trivial.

However, it is important to understand scaling as an epistemological problem. It is not at all clear, as Stevens presupposes, that scales exist “naturally”, independent from observation processes. One could go as far and ask, if a scale does exist at all!

In today’s inter- and cross-disciplinary research, the meaning of scale is therefore at the core of scientific knowledge production (Jones, 1998; Levin, 1992; Manson, 2008; Marceau & Hay, 1999; Marston, 2000). Marston, for example, discusses the use of scales in human geography and argues “for enlarging our scope for understanding scale to include the complex processes of social reproduction and consumption” (Marston, 2000, p. 219). She argues that “space is a social product” (Marston, 2000, p. 221), and refers to Henri Lefebvre (1991) as a “touchstone”. Within this Marxist social-constructivist theory of space, “scale” is not “simply an external fact awaiting discovery”, and

“involves attention to relationships between space and power, and to conceptions and ideologies of space and power that social actors bring to practical efforts to change the world and, of course, to resist change” (Delaney & Leitner, 1997, p. 94 and 96).

With Steven’s positivist account and Marston’s example of Marxist social-constructivism, we have two very divergent concepts of scale at stake. But how do we compensate that contradiction in the conceptualization of scale? Manson offers an “epistemological scale continuum” as a useful rubric for complex human-environment systems

“for reconciling scale concepts and expanding the explanatory power of scale. This continuum ranges from the realist contention that there are natural scales independent of observers to the constructionist argument that social forces actively manipulate scale and its underlying material basis” (Manson, 2008, p. 776).

He differentiates between three types of scale-epistemologies, (1) realist perspectives, (2) hierarchical, and (3) constructionist perspectives on scale, which can be ordered on an interval scale.

“One pole is anchored by the realist ontological premise that there is a single shared reality and the related epistemological claim that reality is readily accessible to objective observers. At the other pole is the constructionist ontological claim that while there may be a reality, in epistemological terms, knowledge about this reality is socially mediated and manipulated. ... Hierarchical perspectives on scale highlight that the observer is critical to defining scale and, in the case of complex emergent hierarchies, demonstrate the potential for subjectivity in interpreting the effects of scale. Hierarchical scales can be reconciled with realist concepts of scale variance, scale invariance, and scale dependence. ... The concept of relative scale essentially mandates the coexistence of multiple observational and explanatory scales that are heavi-

ly conditioned, but not determined, by a shared reality” (Manson, 2008, p. 777 and 782).

What is at stake in the field of political geography, does very well apply to the general scope of “scaling”, too. The broad range of epistemologies described by Manson is also relevant for the other sciences, especially the social sciences⁵.

In addition to the philosophical claims of the “relativity” or “neutrality” of scales, one could argue that the exclusive numerical attribution of information, which is measurable on O- and I-scales, does not quite capture any of the semantically inter-

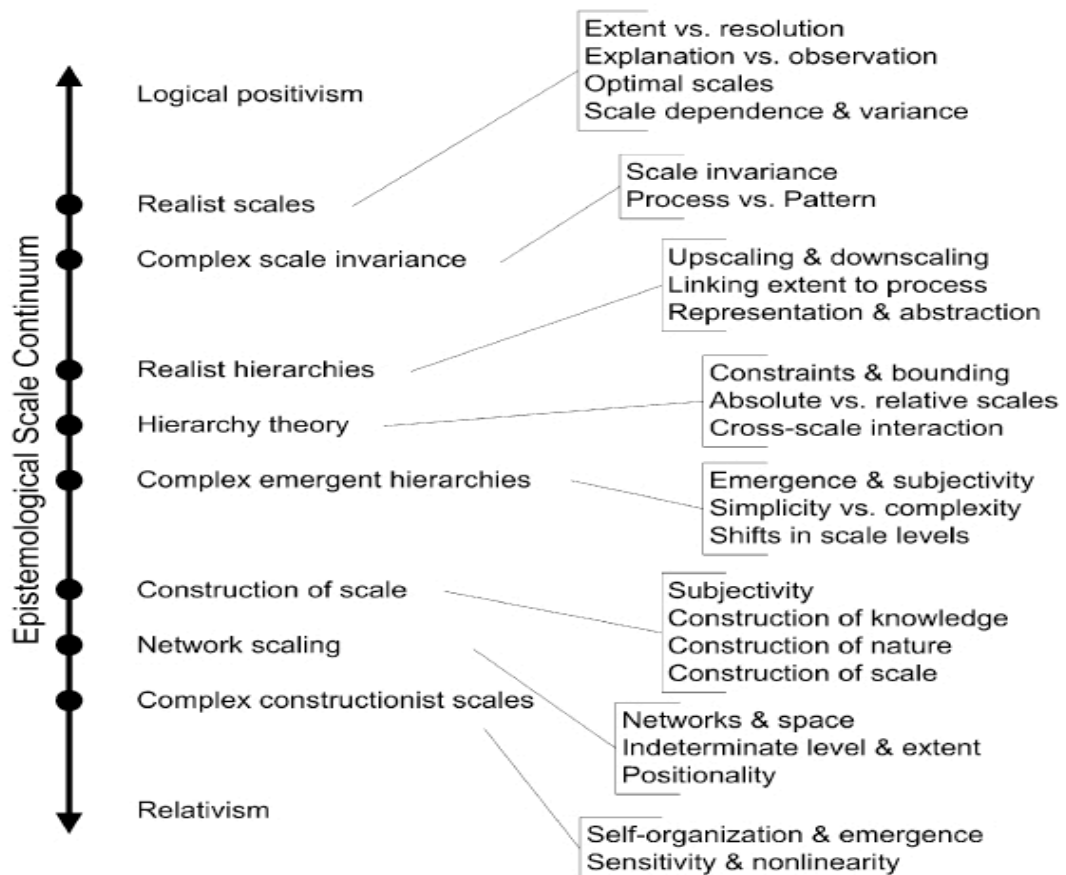


Figure 3. Manson's epistemological scale continuum (Manson, 2008, p. 777).

esting data. Within the “qualitative” paradigm, there is a shift from the universal to the individualistic view (cf. Mjøset, 2000). We are not only interested in the matter of fact that a person is male (coded as “1”) or female (coded as “2”) on an N-scale, or if that person prefers hiking, coded as “1”, to skiing, coded as “2”, or rock-climbing, coded as “3” on an O-scale (see above, 2.1.1). We might learn something about some specific qualities that can be attributed to her hobby, or what hiking etc. means to

⁵ For a further discussion of this topic, please refer to the German „Positivismusstreit“ in the social sciences (Adorno, 1976a, 1976b; Dreier, 1960; Topitsch & Payer, 1993; Weber & Winkelmann, 1951).

her. By “measuring” such sort of information, for example in interviews, video-sequence-analyses, or ethnographical field observations, the researcher re-constructs semantic patterns. Those are categorized under nominally scaled “codes”, and usually, no other operations than the determination of family similarities in the Wittgensteinian sense, (cf. Glock, 1996, "Familienähnlichkeit"), or more precisely, equality-relationships between two or more patterns, are performed.

2.3.2. Tying up the ends of theories of understanding

As Vossenkuhl (1998, p. 170) states, we cannot understand “understanding” without understanding the “either/or-quality” of language and truth. He offers a continuous scale of epistemological theories with Davidson’s (1984) analytical inquiries into truth and radical interpretation at the one, and Gadamer’s (1965) transcendental circular hermeneutics at the other end. According to Davidson (Davidson, 1984, pp. 125-139; Davidson & Harman, 1970, p. 19), “understanding” generates logically from the truth-values of propositions: The proposition “snow is white” for example is only true when snow is white, to quote the most prominent illustration of his truth-conditional theory. So, an adequate theory of meaning according to Davidson should guide us into construing a “systematic theory of meaning *for*” any given language by chronicling the sentence’s composition out of its constituent words, rather than developing a *general* theory of meaning (as Gadamer does at the other end of Vossenkuhl’s line):

“Since there seems to be no clear limit to the number of meaningful expressions, a workable theory must account for the meaning of each expression on the basis of the patterned exhibition of a finite number of features” (Davidson & Harman, 1970, p. 18).

Gadamer, on the other side, refers to dubious metaphysical presumptions along the concept of “truth of being” (German: “Wahrheit des Seins”), which is seen as a vivid quality in every human and makes “understanding” ontologically understandable. But thus his theory cannot explain “*mis*understanding”, which is a common feature in everyday conversation, and is therefore strongly contra-intuitive (Vossenkuhl, 1998, p. 173). So neither Gadamer’s nor Davidson’s concept can exclusively serve as reference theories in a contemporary account of “understanding”.

Oevermann offers a theory that encompasses Davidson’s analytical pragmatism in Peirce’ tradition and the phenomenological “objectivism” in the hermeneutic tradition (Liebau, 1987, pp. 101-107). To remain in Vossenkuhl’s metaphor of an epistemological “line”, one could say that Oevermann ties together the ends of this line and creates thus a circle and adds a second spacial dimension to the metaphor. This comes close to the later Wittgenstein’s account of rule-following with his concept of “tacit knowledge”, at least in the German “transcendental” interpretation of Wittgenstein’s philosophy of language following Habermas (1984) and Apel (1998d). But this transcendental approach is rather ambiguous and not free of interpretational arrogance, as Reichertz (1994, p. 125ff) states, and neither is Oevermann’s account totally consistent nor delivers it a recipe for any given text interpretation, as Oevermann admits himself (Oevermann, 1979a, p. 391f). At this point I want to follow Rorty’s critique of the idea that “objective” truth can be found “out there” (Rorty, 1989, p. 3), which seems the underlying presumption in Oervmann’s concept when

he recurs to Lévi-Strauss' structuralism (Skjervheim, 1982) – especially in the “mirror of nature” (Rorty, 1980).

Rorty (1991b) suggests to dismiss universal claims, such as that “every” language game is contingent (Davidson) or that there exists a language-scheme through which understanding is generated (Apel). Instead, he offers a third way of understanding “understanding” which could be called “practical holism” (Mayer, 1997, p. 198). Here, independent and arbitrary vocabularies exist parallel to each other and “understanding” is generated in a “group united by mutual interests in achieving a common end” and where “conversation is routine inquiry ... by willing to pick up the jargon of the interlocutor rather than translating it into one's own” (Rorty, 1980, p. 318). Where Apel suggests to make an “unavoidable claim to universality” which can be understood only in the analysis of the function of the philosophical language game in strict reflection “upon what one does and presupposes as a philosopher, with the description of particular language games and life forms” (Apel, 1998d, p. 147), Rorty introduces his “principal imperative of solidarity” (Rorty, 1989) as constitutive for understanding. We need to endorse an attitude in which the sense of human solidarity is “a matter of identification with the details of others' lives, rather than a recognition of something antecedently shared” (p. 190). I find this notion much more sympathetic, less imperialistic, and metaphysically much more robust than Oevermann's “objective” approach of translating social observations into his own jargon. And yet – to pick up the above-mentioned metaphor of epistemological theories again – it ties an even tighter knot into the two ends of the line, because it is not grounded in metaphysics but implicit in pragmatic contexts of understanding.

However, Rorty's radical philosophical scepticism needs to be challenged if we intend to work empirically. We need to recur to some sort of accessible “nature.”

2.3.3. Holistic semantics, interpretation and abductive inferentialism

In his “Philosophical Investigations”, Wittgenstein recurs to the metaphor of overlapping semantic “fibres” that are twisted into each other and generate thus a continuity of understanding in a language game that is embedded in the actual context of life forms:

“And we extend our concept of number [“meaning”, u.d.] as in spinning a thread we twist fibre on fibre. And the strength of the thread does not reside in the fact that some one fibre runs through its whole length, but in the overlapping of many fibres. But if someone wished to say: “There is something common to all these constructions—namely the disjunction of all their common properties”—I should reply: Now you are only playing with words. One might as well say: “Something runs through the whole thread— namely the continuous overlapping of those fibres” (Wittgenstein, 1967, § 64).

We must thus look carefully into the milieu of the semantic content in the process of understanding, and at its latent structures. But this means that we need to always *conject* meaning from the context, which is a more sophisticated word for “guessing” (Peirce, 1955), in order to understand something at all, being appreciative to the imperative of solidarity. Finally, interpretation is then a matter of personal attitude dependent from the motivation to understand something/the other.

For the empirical field researcher, this is an important normatively binding obligation for his or her research protocol, and is pre-eminently realised in ethnography/participating observation where the researcher learns about those structures not only cognitively, theory-guided, but also in a pre-verbal, embodied, socially-interactive physical manner, in a “manipulative abductive inference” (Magnani, 2001, cf. 1.2.).

The process of interpretation begins, more than it is acknowledged in classical research in the natural sciences⁶, with the very moment of data collection. If, for the sake of objectivity, “quantitatively” oriented researchers take a distant stance and present their findings “neutrally”, in order to allow others to draw their own conclusions, researchers interested in qualitative features share the conjecture that “neutral” or “objective” data do simply not exist. Each measurement, be it “quantitatively” or “qualitatively” scaled, is semantically pre-determined by the very scope of measurement and research hypotheses with “clear cut problems” (Tukey, 1986 (1961))

So, “quantitative” research designs must not walk right into the trap of presenting “neutral”, i.e. “interpretation-free” data. “Qualitative” research must, on the other side, address the legitimate suspicion to the “prophets of interpretation” who show “virtuosity without scrutiny” (Oevermann, 1979b, p. 390, translated from German by U.D.) by providing information about each interpretational step and each referring theory implicitly or explicitly applied in the interpretation, and leaving also mathematical techniques in “their place”.

Hereby, the interpreter “does not concede complete authorship of meaning structures to the participants in interaction” (Maiwald, 2005, p. 21). The qualitative researcher “conjects” – a *terminus technicus* from literature studies that involves abductive inferring (Wirth, 2008) – meaning from the context, both synchronically (looking at the immediate situation) and diachronically (looking at similar situations). Bateson writes:

„The very possibility of abduction is a little uncanny, and the phenomenon is enormously more widespread than he or she might, at first thought, have supposed. Metaphor, dream, parable, allegory, the whole of art, the whole of science, the whole of religion, the whole of poetry, totemism... the organization of facts in comparative anatomy-all these are instances of abduction, within the human mental sphere. ... Any change in our epistemology will involve shifting our whole system of abductions. We must pass through the threat of that chaos where thought becomes impossible. Every abduction may be seen as a double or multiple description of some object or event or sequence“ (Bateson, 1980, p. 157).

As Bateson describes with reference to Peirce, the epistemic value of “qualitative” empirical research is “real vagueness” respectively “grounded possibility”. Hence, “reality exists of facts and matters of facts and of real possibilities”

⁶ Of course, according to Heisenberg’s uncertainty principle, all scientific measurement is dependent on the experimenter. But in classical, non-relativistic systems, the uncertainty effect induced by the experiment itself is very small in relation to the effect of the system under measurement (cf. Messer, 2007, p. 12).

(Wernecke, 2007, p. 296, translated by U.D.). In “qualitative” empirical research, we must therefore not only develop hypotheses about the factual world by means of induction and deduction, but also aim at propositions about the stochastic world of possibilities. In this abductive step, the interpreter will seek for “real possibilities” of the empirical text-fragments in the tradition of Peirce’s “logic of vagueness”.

“Abduction is here to be seen as an epistemic strategy of innovation as well as an esthetical operation, that is placed in the area of tension between association, imagination, practical reason, and wit” (Wirth, 2003, p. 591, translated by U.D.).

With respect to their epistemic value, propositions that are derived from such an abductive methodology are not completely arbitrary, but “ontologically pre-structured possibilities with respect to an implicit realization” (Wernecke, 2007, p. 296).

However, also within that constructivist epistemology, “quantitative” statistical methods, both, in the semantic analysis of the texts and the exploration of correlations between core categories, can be applied and thus “qualitative” information be scaled by intervals or ratio, too (cf. Falco & García-Lapresta, 2014; Rössler, 2010).

2.3.4. Ontological presumptions of the world of meaning(s) and values

Since the general understanding of today’s qualitative research is that “we do not have access to the world independent from interpretation” (Lenk & Maring, 1997, p. 210, translated from German by U.D.), Lenk and Maring describe in their “fundamental justification of qualitative empirical research”, both, the conceived and the conceivable versions of the world as “constructed and structured” through our “needs, capacities, and possibilities.”

“The world is only conceivable, insofar it (more precisely: its *interpretatum*) has been constructed, structured, and formed by anthropogenic or genetic schemes of interpretation. Everything, that we as perceiving and acting human beings can discern and express, is dependent on schemes of interpretation. ... One could call such philosophy a *transcendental (scheme-)interpretationalism*” (Lenk & Maring, 1997, p. 211, translated from German by U.D., cursive in the original)

Referring to his main publication Lenk (1993), he sees his interpretational, or “conceptual” scheme in the philosophical tradition of Kant, who holds that conceptual schemes, i.e. certain basic concepts, namely the categories of time and space, are indispensable for any intelligible experience of objects at all.

Karl-Otto Apel offers such a pragmatist account within the Kantian respectively Hegelian transcendental heritage – also known as “transcendental pragmatism” (Apel & Papastephanou, 1998). Apel sees language as a *medium*, which reveals truth in the process of communication. In that respect, he is an idealistic verificationist, but he is also a pragmatist, because he endorses an anti-representationalist theory of truth. For Apel, it is the Peircean triadic structure of the sign-relation or semiosis that opens up the realm of understanding.

“This means that, in contradistinction to formal semantics, it does not abstract from the place and function of the sign interpreter - or in traditional terms - of the subject of cognition and his or her truth claims” (Apel, 1998c, p. 73).

But

“Peirce's theory shows that this function has to be integrated into that of an indefinite community of interpretation and consensus formation about truth claims because non-mediated cognition depends on a semiotic process of sign interpretation through 'interpretants' that is indefinite in principle” (ibid.).

This is what makes his account a “transcendental pragmatism” and Apel sees it as a new *prima philosophia* claiming also the later Wittgenstein as his warrant to reformulate the Kantian transcendental *cogito* in a philosophy of language (Apel, 1998d).

Apel thinks that

“Wittgenstein's *Tractatus* should not be understood only as a source of inspiration for logical positivism but also as a kind of transcendental philosophy, namely as a transformation of Kant's critique of pure reason into a 'critique of pure language'. ... In this transformation of Kant's approach the 'supreme principle of synthetic judgments' (that the conditions of the possibility of experience are at the same time the conditions of the possibility of the objects of experience) was to be replaced by the principle that the conditions of the possibility of describing facts through sentences are at the same time the conditions of the possibility of the facts themselves as 'existing states of affairs' (bestehende Sachverhalte) for us” (Apel, 1998b, p. 10).

Obviously, this way of reading the “Tractatus” is an *ex post* interpretation inspired by fragments of the very late Wittgenstein. With “great satisfaction”, Apel quotes an entry from 1931 in Wittgenstein’s “Vermischte Bemerkungen” (1984b, p. 463f), where Wittgenstein refers to Kant’s famous argument of the transcendence of human understanding within the limits of pure reason in § 59 of his “Prolegomena” (Kant, 1957):

“Die Grenze der Sprache zeigt sich in der Unmöglichkeit, die Tatsache zu beschreiben, die einem Satz entspricht..., ohne eben den Satz zu wiederholen. (Wir haben es hier mit der Kantischen Lösung des Problems der Philosophie zu tun) (The limit of language shows itself by the impossibility of describing the fact that corresponds to a sentence..., without repeating precisely that sentence. (We are dealing here with the Kantian solution of the problem of philosophy.))” (Apel, 1998b, p. 10).

Analogously to the Kantian transcendental *cogito*, Apel formulates a transcendental *lego* and understands thus language as ontologically prevalent to the world, which can be seen as the *linguistic* transformation of the Kantian *mentalist* scheme.

However, Lenk argues, that one need not go as far as to buy fully into the transcendental, mentalist part of the Kantian epistemology, and that it might suffice for the moment being, to look at the “method of interpretation” instead and rather speak of a “methodological interpretationalism” referring to “functionalistic” theories of meaning. As Apel, Lenk quotes also a passage from the “Tractatus” (Wittgenstein, 1984a)

to put his *functionalistic* argument forward against “representationalist” theories of truth. A similar approach has been taken by Hilary Putnam, which has become known as “internal realism” (Putnam, 1981, 1987, 1988; Putnam & Conant, 1990). Putnam holds, also in the Wittgensteinian tradition, a pragmatist position, which is, in many points, both anti-realistic and anti-idealist. His “internal realism” claims that, although the world may be *causally* independent of the human mind, the structure of the world - its division into kinds, individuals and categories - is a function of the human mind, and hence the world is not *ontologically* independent (cf. Brown, 1988). Similar to Apel, but somehow more cautious with respect to metaphysical assumptions, Putnam also refers to Kant's concept of the dependence of our knowledge of the world on the “categories of thought” and he claims that there is “*a fact of the matter* as to whether the statements people make are warranted or not” (Putnam, 1981, p. 21, cursive by U.D.). This material, realistic reference allows Putnam to talk about warranted truth that is “independent of whether the majority of one's cultural peers would say it is warranted or unwarranted” (ibid). In this respect, Putnam is more than a mere consensus theorist, but not yet a naturalistic realist. He argues instead that “reason can't be naturalized” (Putnam, 1983) and that here and now “truth is independent of justification ..., but not independent of *all* justification. To claim a statement is true is to claim it could be justified” (Putnam, 1981, p. 56).

It is obvious that language, or rather, the “expressive power of language” (German: “Darstellungsmöglichkeiten der Sprache”, Lenk & Maring, 1997, p. 211), determines the structure of the interpretational scheme, and that language is the key to all qualitative empirical research.

Accordingly, Atkinson sees “sociology is a rhetorical activity” (Atkinson, 1990, p. 10) and in any observation process we reconstruct “the world as text” (Garz & Kraimer, 1994, p. 7). Language is

“the most common form of meaningful expression, and even if the researcher is looking at images or observing social practices, these are commonly transformed into linguistic form through descriptions and field notes. ... Language not only incorporates the terminology and vocabulary with which we understand the world, use it, transform it, but also is the medium by which we convey that meaning or interpretation to others” (Gibbs, 2002, p. 1).

The waves of the “Objektivismusstreit” from the late 60ies seem to have calmed in the 80ies, the linguistic turn has been widely accepted in social sciences (Skjervheim, 1982), and philosophical models have been developed in order to render realism thinkable after the linguistic turn (Habermas, 2001).

However, the epistemological and ontological presumptions of the “world of meaning(s) are – irrespective of the pragmatic truce – not at all that clear, and so-called “hermeneutical” methodology is still rather obscure in today's social science and “a battleground of competing opinions” (Bernstein, 2011, p. 15).

The main challenge of hermeneutical methodology to empirical science is the implicit refutation of the claim, that the world is accessible independently from the inter-

pretation through our senses and language⁷, which we have formulated as the first default position of positivist empirical science. It is thus conceptually impossible within the “qualitative” paradigm, to separate data “analysis” from data “collection” in the field.

Some authors go as far as to avoid the terminology of “analysis” at all. Silverman (2011), uses the term “interpreting” instead, Wolcott (1994) prefers “transforming”, and Hammersley and Atkinson (2007) refer to “making sense of” the data.

So, in contrast to that, we can frame as a first default position of hermeneutical methodology:

- There might be a real world that exists independently of us, but it is not accessible for us independently of our experiences, our thoughts, our language.

The weaker interpretation of this claim, i.e. that the world may be causally but not ontologically independent of the human mind, what we have called “methodological interpretationalism”, could still be shared by the positivist empiricists, as we have seen in Vollmer’s projective epistemology (cf. **Figure 1**).

However, in the more radical and metaphysically less presuppositional understanding, this representationalist and mentalist epistemology is seen as “indefensible” (Habermas, 1992, p. 149). In this 1992-essay, Habermas discusses Searle’s analyses of communication, which has become known as “speech acts theory” in the history of philosophy (Searle, 1969) and confronts Searle’s theory with his non-representationalist/non-correspondence-theoretical pragmatism which itself is constitutive of most of today’s hermeneutical concepts in social science (Bernstein, 2011). According to Habermas, such mentalist concepts gain their epistemic power from “the myth of the given”. But after the linguistic turn, direct access to “some inner or outer reality” is philosophically denied and

“the alleged immediacy of sense-data cannot any longer function as the ultimate and infallible appellate instance. Without the possibility of recourse to the un-interpreted sense-material, sense-data lose their unquestioned authority” (Habermas, 1999b, p. 20, translated from German by U.D.).

Accordingly, Searle’s second default position must be revised as follows:

- We have *no* direct perceptual access to the world through our senses.

The word “typical” in Searle’s third default position (see above) refers to his “sharp and complex analyses” (Habermas, 1992, p. 149, translated from German by U.D.) and shows that the role of language cannot be unequivocally or stereotypically be attributed to either epistemological tradition. However, default positions four and

⁷ Depending on the radicalism of the interpreter, this might or might not include the assumption that the world *exists* independently of our perception. But we can well leave this question to the philosophers.

five, the referrals to a correspondence theory of truth and causal ontology, are clearly out-of-line in critical hermeneutics.

As we have seen above in his critique on Searle, Habermas is a little bit closer to the classical “analytical” pragmatism than Apel’s “transcendental pragmatism” or Lenk’s “transcendental interpretationalism”, respectively. As Putnam, Habermas is much more cautious about transcendence and formulating a metaphysically weaker position. His theory of communicative action secures understanding only in a transitory semantic process. Ideally, this transition is a domination-free act of communication in democratic institutions. Truth evolves from a “consensus” in such a society, but it is a *pragmatically constructed* truth of a world that is intermediated by our mind - what is called “weak naturalism” by Habermas with reference to Kant and Darwin (Habermas, 1999b, p. 32ff).

Habermas argues, with reference to Rorty (1980), that the “mirror of nature”, i.e. the representation of reality, is the wrong epistemological model, because the two-dimensional relation of “image” and “screen” as well as the static relation between “matter of fact” and “proposition”, ignores the dynamics of the growth of knowledge through problem-solving and justification. Instead of endorsing that static representationalist epistemology, Habermas sees the main thrust of any cognitive process in the “intelligent reaction to a risky world”, in pragmatic problem-solving and performative action. Hereby, Habermas maintains the transcendental Kantian distinction of the “world” on the one side, that is seen as a “totality of objects (rather than ‘matter of facts’) for possible references”, and some “inner” self, which is in principle and with reference to Darwin’s (and Vollmer’s) evolutionary epistemology, neurologically or bio-genetically explainable, and which is reacting upon some objectively given environment. Hereby, the inner-perspective of the subject and the outer-perspective of the objective world must not – in explicit juxtaposition to Vollmer – be further ontologically defined. Such definition would inevitably lead to infeasible metaphysical assumptions and reduce culture to nature, whereas Habermas sees continuity on a meta-theoretical level. Our socio-cultural life-forms are adaptations to the world, unremitting to our biological adaptations in the course of the evolution of *homo sapiens* (cf. Habermas, 1999b, p. 36ff). Habermas acknowledges that this is a somewhat “unusual” argumentation, which overtly refrains from any philosophical assumption about the relation of body and mind. Accordingly, Dews characterizes this ontological claim – with Habermas’ *placet* (cf. Habermas, 1999b, pp. 39, footnote 41) – as a combination of anti-idealism with anti-scientism and a propensity toward naturalism (cf. Dews, 2001), or as a manifestation of manipulative abduction (Magnani, 2001).

2.3.5. Leaving the methodological dualism behind: Approaching the “Third Paradigm”

From an empirical perspective, this Habermasian epistemological stance seems to be the perfect starting point for research “in the mirror of nature”. Accordingly, Johnson and Onwuegbuzie (2004) position a third paradigm, *mixed method research*,

“as a natural complement to traditional qualitative and quantitative research [and] present pragmatism as offering an attractive philosophical partner for mixed methods research, and to provide a framework for designing and conducting missed methods research” (Johnson & Onwuegbuzie, 2004, p. 14).

Johnson and Onwuegbuzie discuss the strengths and the weaknesses of each of the three research paradigms and offer a systematic analysis of different research strategies. The weaknesses debated for pure “quantitative research” resemble very much the paradigmatic discussion of Fenger’s dissertation (cf. 2.2), i.e.

- that “the researcher’s categories [...] may not reflect local constituencies’ understandings”,
- that “the researcher’s theories [...] may not reflect local constituencies’ understandings”,
- that “the researcher may miss out on phenomena occurring because of the focus on theory or hypothesis testing rather than on theory or hypothesis generation (called the confirmation bias)”,
- that “knowledge produced may be too abstract and general for direct application to specific local situations, contexts, and individuals” (Johnson & Onwuegbuzie, 2004, p. 19).

The general weaknesses associated with pure “qualitative research” include:

- “Knowledge produced may not generalize to other people or other settings (i.e., findings may be unique to the relatively few people included in the research study).
- It is difficult to make quantitative predictions.
- It is more difficult to test hypotheses and theories.
- It may have lower credibility with some administrators and commissioners of programs.
- It generally takes more time to collect the data when compared to quantitative research.
- Data analysis is often time consuming.
- The results are more easily influenced by the researcher’s personal biases and idiosyncrasies” (Johnson & Onwuegbuzie, 2004, p. 20).

Now, mixed methods research overcomes those weaknesses since

- “Words, pictures, and narrative can be used to add meaning to numbers,
- numbers can be used to add precision to words, pictures, and narrative.”

Furthermore, mixed methods

- “can provide quantitative and qualitative research strengths, [...]”
- can answer a broader and more complete range of research questions because the researcher is not confined to a single method or approach.
- A researcher can use the strengths of an additional method to overcome the weaknesses in another method by using both in a research study.

- Can provide stronger evidence for a conclusion through convergence and corroboration of findings” (Johnson & Onwuegbuzie, 2004, p. 21).

This becomes even clearer if we look at the above-presented epistemologies underlying “quantitative”, respectively “qualitative” research. We can see two interesting features:

The underlying ontological assumptions, formulated in Searle’s commonsensical “default positions” become more and more contra-intuitive and less understandable with the attenuation of their metaphysical ingredients. In other words: despite the anti-transcendentalist claim of the positivist sciences, this form of sciences derives upon much more substantial metaphysical assumptions than so-called constructivist “qualitative” methodologies. But with increasing scepticism ceases comprehensibility and commonsensical acceptability of science.

However, the stronger metaphysical claim made by the (more) positivist epistemologies entails the weaker positions as *possibilities*. And the weaker, i.e. (more) sceptical positions, can refer to the (more) positivist accounts as the possible end-points of their pragmatic position. Within Habermas’ or Putnam’s pragmatist philosophical framework, all of the above-presented epistemologies can be outlined on a continuous scale and each is suggesting its own epistemic power at the cost of metaphysical assumptions (cf. Table 5), which become – in a sort of performative self-contradiction – philosophically unwarranted if exclusively postulated, since such ultimate foundation would require transcendental arguments which is a counter-indicated presupposition (Apel, 1998a).

In the following, I will briefly present the mixed-method strategy of the two empirical articles attached in the appendix, and I will discuss the underlying epistemological and metaphysical assumptions according to Alf Odden’s classification (Odden, 2001).

	Theory of Truth	Ontological Subject-Object Relation	Epistemic Operation	Epistemic Strategy	Epistemic Power*	Metaphysical Claim	Metaphysical Warranty
Naturalism (Vollmer, 1994)	representation on a biological "screen"	static projection	mentalist	positivistic	very strong	very strong	very weak**
Transcendental Interpretationalism/ Pragmatism (Apel & Papastephanou, 1998; Lenk, 1993)	representation in a conceptual scheme	dynamic interpretational apperception	mentalist/ linguistic	constructivist	strong	strong	weak
Methodological Interpretationalism/ Internal Realism (Lenk & Maring, 1997; Putnam, 1981, 1983, 1987, 1988)	functionalism	dynamic interpretational reaction to the world	pragmatist/ linguistic	constructivist	weak	weak	strong
Weak Naturalism (Habermas, 1992, 1995, 1999a)	justification/ discourse	dynamic performative reaction to the world	pragmatist/ linguistic	constructivist	weak	weak	strong
Radical Interpretationalism (Davidson, 1984; Davidson & Harman, 1970; Rorty, 1989, 1991a, 1991b)	interpretation	Irony, principal of charity	pre-verbal and linguistic	constructivist	very weak***	very weak	very strong

* According to commonsensical epistemology (Searle, 1999).

** Unwarranted if claimed as the exclusive epistemology (Apel, 1998a; Habermas, 1999b).

***Non-sensical in empirical research.

Table 5. Comparison of epistemological and metaphysical features of different accounts of “realism” and “naturalism”. As has been argued in the previous chapter, some sort of “weak naturalism” or “internal realism” seems rational for mixed-method empirical research in the mirror of nature.

3. Mixed-method research in the mirror of nature: Two empirical surveys within the "Third Paradigm"

3.1. Alien at home...

3.1.1. Summary, research problem and theoretical model

In „Alien at home...“, we searched into the “adjustment strategies of students returning from a six-months over-sea’s educational programme” (Dettweiler, Ünlü, Lauterbach, Legl, et al., 2015). In a second step, we tested our data against the UU-curve theory of Reverse Culture Shock (Gaw, 2000) or Expedition Reverse Culture Shock (Allison et al., 2011), respectively. In order to address the above-mentioned flaws of mono-method research designs (cf. 2.3.5), we had chosen a mixed-method design.

This invoked a few difficulties:

First, we had to approach the research problem as a team in order to achieve a high coding reliability with two independent coding-strategies performed by two coders. The consequence was a rather time-consuming and expensive study, which could be partially financed by a grant of the Staedtler-Stiftung to the University of Erlangen and an anonymous 40.000 Euro donation to TUM from the parents of one “KUSi”. Furthermore, I, as the principle investigator, had to “learn about multiple methods and approaches” and had to “understand how to mix them appropriately” (Johnson & Onwuegbuzie, 2004, p. 21).

Generally speaking, we chose a constructivist epistemology, along the lines of “understanding”, which is usually subsumed under “qualitative research methodology”. However, within that constructivist epistemology, we apply “quantitative” statistical methods, both, in the semantic analysis of the texts and the exploration of data distribution patterns in the core categories. In Johnson’s and Onwuegbuzie’s classification of mixed-method research, this approach is

qual → *Quan*.

“Qual” stands for “qualitative”, “quan” for “quantitative”. The arrow “→” stands for sequential (as opposed to “+”, concurrent). Capital letters denote high priority or weight, lower case letters lower priority or weight (Johnson & Onwuegbuzie, 2004, p. 22).

3.1.2. Methods of data collection and empirical model

In order to achieve a critical grade of comparability to Allison, Davis-Berman, et al. (2011) findings, we chose a similar methodological approach in the data collection. We sent out letters to all participants of all four KUS-cruises and asked them to write

us back telling us about their experiences after they had returned home. Each subject received a personal note (N=128) explaining the research goal, a line-numbered, pre-coded (gender and cruise number) template to write on, and a ready-stamped return envelope. Anonymity and independence of the study from the "KUS-Projekt" were ensured. The hand written letters had been transcribed into a word-processing software and additionally been digitalized with pdf-scans.

We received back 56 letters ($n_1 = 6$; $n_2 = 9$; $n_3 = 20$; $n_4 = 21$) which is a global response rate of $r_g = .44$. The global response rate can be assumed very high, given a standard value of $r_s = .05$ (Porst, 2001) and Allison's $r = .26$, but cruises (1) and (2) are underrepresented ($r_1 = .19$; $r_2 = .28$) with reference to the other two cruises ($r_3 = .63$; $r_4 = .66$) which leads to a bias in the stochastic interpretation and needs some attention in the statistical model. We explain the increase of motivation to participate in the survey and returning their report mainly by the time factor. From the letters we received, the general life circumstances and backgrounds across the four groups are rather constant and we do not have any passages in the texts that hint at other reasons than increasing indifference to the research question with growing remoteness from the experience. The relative jump in the response rate after turn 3, i.e. 8-20 months after re-entry, indicates that a certain distance seems to make quite a difference.

3.1.3. Methods of data interpretation and statistical model

Independent coding of the text with NVivo® live-coding and qualitative content analysis to find core categories

In order to achieve sufficient inter-coder reliability of the qualitative interpretation of the data, two coders have applied independent coding strategies. Coder₁ accessed the digitalized text with NVivo® live-coding (Bazeley, 2007; Bringer, Johnston, & Brackenridge, 2006; Gibbs, 2002; Lewis, 2004), whereas coder₂ performed "classical" content analysis on the hand-written material (Flick, Kardorff, & Steinke, 2004). The coders did not specifically look for culture-shock symptoms, however, since we had asked the students to tell us about their time at home, and how they perceived the re-entry, re-adjustment was a predominant topic. All authors compared the two coded sets, and five core categories had been extracted as a consensus.

Coder₁ produced an NVivo® output on the basis of the core categories, and coder₂ re-coded the hand written material accordingly. Data are originally in German, and all interpretational operations were conducted in German on the German text. Only after the analysis, we have translated those text-chunks that are directly quoted in this paper into English for the sake of international dissemination.

Quantitative cluster analysis with NVivo® and quantitative encoding of the categories

In order to gain an overview of the quantitative features of the text, coder₁ used NVivo® to produce cluster analyses, including word counts, coding densities, and semantic clusters. Additionally, both coders categorized all text samples with respect to the five core categories in an interval scale from -2 to 2 (strong negative manifestation, medium negative manifestation, neutral manifestation, medium positive manifestation, strong positive manifestation) and produced the mean-values for each category for each letter. This method is an adaptation of Rössler's "Inhaltsanalyse" (content analysis) which is designed for communication science (Rössler, 2010) and has recently been described in mathematical terms by Falco and García-Lapresta (2014).

Principal Components Analysis (PCA) across the five categories and their polynomial interpolation (PI)

In order to search for underlying structures in our five categories, we performed Principal Component Analyses (PCA) with the free software package "R". "PCA seeks to composite scores of observed variables" (Kasper & Ünlü, 2013, p. 2) rather than latent variables, and to "reduce the measured variables to a smaller set of composite components that capture as much information as possible in the measured variables with as few components as possible" (Park, Dailey, & Lemus, 2002, p. 563).

We hypothesized that the coding of the categories produced possibly correlated variables, and PCA allows to convert those to a set of values of linearly uncorrelated variables, which are called "principal components" after Hotelling (1933) and carry a loading matrix which can be computed rather than estimated. Hence, PCA is not dependent on a critical number of data and is especially suitable for small sample sizes (Kasper & Ünlü, 2013). Thus, we were able to have a stochastically independent look on each category and account for their distribution patterns across the four cruises by interpolating the points of mean- and median-values in each category into the "underlying" polynomial function and testing our data against the "second" U in the UU (W)-curve theory, which would be mathematically expressed by a 2nd grade polynomial function with $a > 1$.

Development and test-of-fit of the statistical model and stochastic exploration of significant differences of the groups

In order to check for non-normal reference points in variables of the PI, and in order to look into potential gender effects, we compared the mean values of the five categories over the four cruises with one-way ANOVA using Tamhane's T2 *post-hoc* test in SPSS 22, which is designed for situations in which population variances differ and which is rather conservative and robust to small sample sizes (Tamhane, 1979). In order to minimize the level of variance and stabilize the data, we decided to run the ANOVA on a median-split of the data (J. Cohen, 1983).

Qualitative Contextualization

In the last interpretational step, the findings of the quantitative analysis have been contextualized with anchor examples in order to discuss the data in their original form. We chose to concentrate on examining only a few "anchor examples" (Mayring, 2014) in order to exemplify each topic.

3.1.4. Findings and Epistemological considerations

In this paper, we have given empirical support to the (second) U-curve stage model in expedition reverse culture shock theory, meaning that the first phase after arrival back home may appear easy enough but that feelings of psychological strain and isolation are very likely to follow in the course of the first year. Approximately one year after arrival, the support from the peer-group (KUSis) together with personal growth facilitates the re-adjustment to the alien old world, or, in other words, "to the home culture as a new culture" (J. N. Martin, 1984, p. 121).

Furthermore, our findings confirm very strongly the first category that Allison et al. (2011) came up with, namely the "Sense of Isolation". We were also able to discern some of the symptoms Oberg talks about, most of all "psychological strain", "a sense of loss" and "a feeling of deprivation", maybe even as far as feelings of rejection by the new – which actually is the old – culture (Oberg, 2006(1960)). All of our students, boys and girls equally, described very similar feelings about their coming home – sometimes even with almost the same words as the young explorers in Great Britain (Allison et al., 2011). This might suggest that the duration of the journey is not that important, as our youths stayed abroad for half a year whereas the expedition of the British kids lasted only for six weeks. It seems that the remoteness of the places visited, the close companionship with people in the same situation, and the vivid impressions of the journey have a much deeper impact.

Gender does not seem to make a difference which contradicts empirical findings of Brabant, Palmer, and Gramling (1990) who concede that re-entry may be more problematic for females than for males. And we could also not find the opposite, i.e. that "women were more satisfied with re-entry life", as Rohrlich and Martin (1991) found in their study. However, as we have discussed above, our group may not be typical with respect to gender-role variance and the pedagogical intervention may have done its part here, too. Neither were we able to discern different coping styles (Adler, 1981), which may be due to the very homogenous group selected in the application process.

We think that the mix of methodological approaches to the reports pays off the flaws that each individual method might have. The inhomogeneous response-rate with its critical effect to the statistical operation, is paid off by the possibility to speak of "significant" or "non-significant" differences between the groups and reveals effects which would not have been visible by just reading the texts and analysing them by means of "classical content analysis". Most importantly, the principal component analysis (PCA) and the polynomial interpolation revealed patterns that could not be detected by any other of our methods that clearly – but the plots would, of course, be useless without the understanding from the texts themselves. The NVivo®-coding and semantic analysis revealed rather superficial patterns of the texts, and many of our originally performed analyses were simply pointless. However, the word-counts and coding-density-analyses uncovered textual features and patterns that would, without the use of computing techniques, not have been detectable by classical content analysis. Finally, the classical content analysis, which bares the most profound but also the most "subjective" understanding of the texts, is escorted by robust "objective" data. It would have been useful to perform follow-up interviews for in-depth

understanding, e.g. with respect to our hypotheses concerning the camera-team in cruise 3, or the U-curve allegations. Of course, a longitudinal design would have promised much better results. However, man and women-power in this project were limited, and more interviews and prolonging of the data-interpretation beyond the budget.

We nevertheless think to have shown that the returning students show symptoms of expeditionary reverse culture shock, very similar to what Allison et al. have found. We have also shown that there is hardly any difference of perception of the expeditionary experience over time, and that the students are very competent in developing coping strategies within their peer group. And we must not forget that despite all adjustment problems, we found many passages in the texts that express great gratitude for the opportunities offered by the programme, which helped the sojourners grow. This corresponds well with the findings of Kaplan and Talbot (1983) who took students on week-long wilderness trips where students reported positive psychological experiences, such as "increased ability to distinguish the significant from the trivial" (Kaplan & Talbot, 1983, p. 184).

„Alien at home...“ applies clearly the Habermasian “weak naturalism” and all methods used in the design are stringently deduced from this epistemological presupposition, since despite all quantitative statistical analyses and mathematical computations, the general paradigm is qualitative (qual → Quan).

The only critical methodological operation is the ANOVA on the median-split of the data set. Here, we apply “pre-revolutionary NHST-techniques” (Rodgers, 2010). However, (a) our sample size is critically small, (b) descriptive data diagnostics support the analyses, (c) we used robust enough tests, and (d) we can – different from Fengler – contextualize those data with first-hand qualitative information providing thus logical support for the p-values and warrant also practical significance.

3.2. Pupils’ learning motivation during outdoor science teaching

3.2.1. Summary, research problem and theoretical model

In this article, we searched into learning psychological aspects of an outdoor science teaching programme within Self-determination Theory (Deci and Ryan (2000)). As in “Alien at home...”, we made use of a mixed-method approach in the *qual* → *Quan* paradigm (Johnson & Onwuegbuzie, 2004), facing the same logistical challenges.

Moreover, in this study, we had not only to finance the evaluative research, but also the research object, i.e. the “Forscherwochen” at the Schülerforschungszentrum Berchtesgadener Land. Fortunately, the research weeks have been made possible by two grants, grant #2012-057 by Hans Sauer Stiftung München and one by Kiel & Ruder e. V., München, summing up to 30.000 Euro.

Self-determination theory in the pedagogical context proposes that the pupils’ motivational behavior is dependent on the satisfaction of certain psychological needs, i.e. the „opportunities to experience autonomy, competence, and relatedness“ (Levesque, 2004, p. 68).

According to White (1959), competence stands for the demand to control the outcome and experience mastery. Autonomy is the universal urge to be causal agents of one's own life and act in harmony with one's integrated self (Deci & Vansteenkiste, 2004). Relatedness is the universal want to interact, be connected to, and experience caring for others (Deci & Ryan, 1985; Deci & Ryan, 2000; Deci & Ryan, 2002; Deci & Vansteenkiste, 2004) and the better those basic needs are satisfied, the more self-regulated are the pupils’ motivational behavior patterns (Deci & Ryan, 1985; Deci & Ryan, 2000; Deci & Ryan, 2002; Müller et al., 2007).

The learning motivational behaviour is defined on a continuous scale of self-determined action. Deci and Ryan divide this scale in three different types of motivation: intrinsic motivation, extrinsic motivation and amotivation. Both, intrinsic motivation and amotivation are not further differentiated, whereas extrinsic motivation is segmented into four types of regulation, i.e. integrated (InR), identified (IdR), introjected (IjR), and external (ExR) regulation (Deci & Ryan, 2000; Deci & Ryan, 2002).

The main interest in our research design was to understand the complex of learning motivational aspects with certain didactic features of the programme in the context of the outdoor educational setting during the research weeks. Data from a pre-test survey suggested positive effects of the research weeks (intervention) on the pupils’ motivational behavior in science lessons. Accordingly, three guiding hypotheses (HG) could be formulated:

HG₁: The pupils’ learning motivational behaviour measured in the context of the research week (FoWo) is higher than at the classroom (NuT).⁸

HG₂: The pupils perceive the practical orientation of the programme at FoWo as higher than at NuT.

HG₃: Group dynamics and physical activity levels correlate positively with self-regulated motivational behaviour.

Each guiding hypothesis contains a number of sub-hypotheses, which can be referred to in detail in Table 6 .

A forth set of questions searches into cross- and gender-effects between the three different complexes. Therefore, we did not provide a set of a priori defined hypotheses but rather chose a pragmatist constructivist epistemology along the lines of “understanding phenomena” a posteriori (Biesta & Burbules, 2003).

⁸ “FoWo” is short for German “Forscherwoche”, “research week” (outdoor). “NuT” is short for German “Natur und Technik”, “nature and technology” (indoor).

3.2.2. Methods of data collection and empirical model

Each of the three above-mentioned hypotheses requires different methodological approaches summing up to a mixed-method approach. Data for the thematic complex of learning motivational behaviour is collected by ethnographic field-notes from pupil observation (Bogner, Emerson, Fretz, & Shaw, 2002; Emerson, 2001; Hammersley & Atkinson, 2007; Koepping, 1987) and pupil questioning including open ended questionnaires as well as five interrogations conducted during one intervention as a group-interview (Atkinson, 1990; Flick et al., 2004). Hereby, student observation was performed on a general group level. For ethical reasons, the observed behavior has been documented anonymously. For the same reason, we did not identify the children in the group-interview, who had been randomly chosen, three girls and two boys, with their questionnaires. Those, however, had been pre-coded and personalized in a pseudo-anonymous way, so that the person looking at the questionnaires in the school- and in the outdoor-education contexts could issue the corresponding pairs of questionnaires. Data from the questionnaires are quoted in the essay with the syntax “#school, week, student number, gender”.

Quantitative measures with the German version of the Academic Self-Regulation Questionnaire (SRQ-A) originally developed by Ryan and Connell (1989), and adopted especially for our target group by Müller et al. (2007) complete the picture we aimed to gain from the pupils’ learning motivational behaviour. The SDI is calculated from the abovementioned four motivational domains, intrinsic motivation (InR), identified motivation (IdR), introjected motivation (IjR), and extrinsic motivation (ExR):

$$SDI=(2 \cdot InR+IdR) - (IjR+2 \cdot ExR).$$

The second complex of themes, practical orientation of the didactic concept, was searched into by issuing a questionnaire consisting of five items, each indicating a different type of practical relevance. This inventory was adopted from the scale “practical orientation” developed by Rakoczy et al. (2005) and is validated especially for our target group.

Data on complex three, group coherence and factors of well-being together with physical activity level, were gained by means of pupil observation and pupil questioning, using both, open (qualitative) and closed inquiries (ordinally scaled data). Quantitative features of the pupils’ activity level have been determined by the recording of the pupils’ heart rates with acentasTM heart rate monitoring devices. However, only a short but representative sample of 6:30 hours, taken from two cohorts in school B, has been included in the analysis.

We tested n=84 pupils in three different weeks in 2013 (24; 26; 39) from two different schools, referred to as A (n=20) and B (n=64). All pupils were between 10-12 years old, and consisted of 43 girls and 41 boys. The schools differ with respect to their academic profiles. Whereas school B is a university reference school known for its science teaching, school A is a so-called “Elite School of Sports” accredited by the German Olympic Sports Confederation, and some of the pupils in school A were

prospective elite athletes. Consequently, the pupils’ behaviour and academic performance throughout the intervention was assumedly different.

3.2.3. Methods of data interpretation and statistical model

Data analysis was conducted according to the requisites of the given methods of data collection, including statistical analyses of quantified variables from the questionnaires, quantitative linguistic analyses, and classical content analyses for the data from interviews, field notes, and open questionnaires. As in “Alien at home...”, we warranted sufficient inter-coder reliability of the qualitative interpretation of the data by two coders applying independent coding strategies. Coder₁ accessed the digitalized text with NVivo® live-coding (Bazeley, 2007; Bringer et al., 2006; Gibbs, 2002; Lewis, 2004), whereas coder₂ performed “classical” content analysis on the hand-written material (Flick, 2007; Flick et al., 2004; Mayring, 2014). All authors compared the two coded sets, and the categories had been extracted as a consensus.

Coder₁ produced an NVivo® output on the basis of the core categories, and coder₂ re-coded the hand written material accordingly. Data are originally in German, and all interpretational operations were conducted in German on the German texts. Only after the analysis, we have translated those text-chunks that are directly quoted in this paper into English for the sake of international dissemination.

The calculation of simple descriptive statistics, such as means and its graphical visualizations, allowed us a deeper interpretation of the data and indicated which groups could be reasonably compared with respect to p-values. We used interactive graphics to get an overview of central descriptive statistics regarding the relevant variables, and further analyses have been undertaken with Trellis-plots that give graphical visualizations in the form of matrix displays and enable easy and efficient exploration of multivariate data (Cleveland & Becker, 1996; Sarkar, 2008). Mean comparisons and one-way ANOVAs using the more robust Brown-Forsythe and Welch test were applied to identify significances in our datasets (Garson, 2012). We tested the assumptions of one-way ANOVA. The homogeneity of variances were examined by Levene’s test, the normality assumption of the residuals could be attested by skewness, Shapiro-Wilk, and Kolmogorov-Smirnov test, as well as with appropriate diagnostic plots.

To examine the increase of SDI from the school to the outdoor context ($\text{diff}_{\text{SDI}} = \text{SDI}_{\text{FoWo}} - \text{SDI}_{\text{NuT}}$), we utilized also regression techniques, such as linear regression and (orthogonal) polynomial regression. In this regard, we initially explored the monotone linear relationships among the SDI variables (SDI_{NuT} , SDI_{FoWo} , and diff_{SDI}) by a simple scatterplot matrix. Furthermore, we computed a simple linear model, fitting diff_{SDI} values against SDI_{NuT} values, and tested the assumptions of linear regression - that are (a) linearity, (b) normal distribution of residuals (c) homogeneity of variances, and (d) outliers or influential points (leverage; Cook’s distance) - by appropriate diagnostic plots (J. Cohen, 2010).

Reliability of the inventories have been calculated using Cronbach’s alpha (Cronbach, 1951; Cronbach & Shavelson, 2004). Additionally, we performed a contextual analysis of quantitative features of the qualitative data from the pupil questioning with the SDI-measures in order to validate the plausibility of the findings (Mayring, 2014).

In order to search into the effects of the “during-course” factors with respect to learning motivation, we executed correlation analyses. For measuring statistical dependence between the ordinaly scaled items, we used Spearman’s rank correlation coefficient (ρ) and paired t-tests to search into differences in the separate teaching environments, indoor vs. outdoor. For the computations, we used SPSS and the open source software R (www.R-project.org).

3.2.4. Findings and Epistemological considerations

The most obvious finding in this survey is clearly that in the outdoor educational setting, pupils show significantly higher learning motivational behaviour – irrespective of gender or school culture. In terms of relatedness, school culture has a significant influence on the level of self-regulated motivational behaviour. Moreover, less self-regulated pupils profit more from the outdoor setting than those who show already a high intrinsic learning regulation – independent from the relative SDI-level, as the school comparison shows.

This corresponds well with recent comparative didactical research conducted in the context of student laboratories. Thomas and Müller (2014) investigated into differences of regular science classes and science labs, regarding the pupils’ perceived autonomy support, intrinsic motivation, and identified regulation. Their theoretical frame was identical with our survey within SDT, using the same psychometric scale SRQ-A (Müller et al., 2007). They report that in grades five and six perceived autonomy support and intrinsic motivation were high both in science classes and in science labs. At grades seven and eight perceived autonomy support and intrinsic motivation were still high in science labs but not in science classes.

With our data, we can fully support those findings for the high values in science class with the younger age group. We have not yet searched into older pupils and will address this theoretically very interesting problem in our next surveys. However, when we compare the mean-values for intrinsic and identified regulations of boys and girls in the context of student labs and the outdoor educational setting, we can concede that both girls and boys show even a higher intrinsic motivational behaviour in the outdoors than the girls and boys in Thomas and Müller’s survey in the student labs. The same pattern can be detected with identified regulation. In the context of the outdoor teaching, both, girls and boys show higher values than the boys and girls at the research lab. Even if we have not directly measured perceived autonomy, as Thomas and Müller have, we can explain this pattern by theorizing that there is a linear progress of perceived autonomy from class via student labs (mostly one-day interventions) to our research weeks in the outdoors, where there is a lot of time for the pupils to engage in the research tasks and perceive competence and autonomy.

This is well supported by our finding that the outdoor-teaching has been perceived by the pupils as being generally more practical than the teaching at the normal school context. There, “practical” stands indirectly for the two basic needs of autonomy and competence (Rakoczy, Klieme, & Pauli, 2008) and promotes the connection of science (class) to every day life experiences. The latter is hypothesized as “not yet taken fully advantage of” in student labs and science classes by Thomas and Müller (2014, p. 56).

If we look at the pupils’ behaviour during the outdoor-research, and also consider the qualitative feedback we received in the questionnaires, we can easily understand that the pupils enjoyed the consequent practical orientation of the science teaching and express to have had great “fun”, by doing things themselves (e.g. building their own measurement tools and develop their own data-collection strategy). Additionally, they perceived themselves as competent (and got instant feedback, when their self-made measurement tool showed the same humidity as the pro-tool in their backpack) – and these seem to be the main influential factors for the highly self-regulated motivational behaviour. Again, we cannot account for any gender effects with respect to the variables of practical orientation of the programme which accords to findings of Engeln (2004), Brandt (2005), and Guderian (2007), and this “might be a promising starting point for reducing the gender gap in the motivation for science and for winning over young people to careers in the sciences” (Thomas & Müller, 2014, p. 56).

Moreover, in addition to the didactically outstanding work being done at student laboratories, the physical activity in the outdoors (hiking) and the group dynamics in the one-week residential outdoor setting can be seen as factors that contribute to the pupils’ exceptionally high self-regulated learning motivational behaviour. This accords with findings in Denmark, where pupils who are taught outside the classroom show significantly higher activity levels than those who are not (Mygind, 2007), and that the outdoor teaching leads to better social relations and experiences with teachers (Bentsen, Mygind, & Randrup, 2009).

Furthermore, we have seen that the “fun” and “enjoyment” inherent in the hiking and the practical orientation inherent in the scientific research projects add to the already stimulating pedagogical setting. This is in accordance with Bisson and Luckner (1996, p. 108) who searched into the pedagogical role of fun in adventure education and found that “fun can have a positive effect on the learning process by inviting intrinsic motivation, suspending one’s social inhibitions, reducing stress, and creating a state of relaxed alertness.”

In addition to the research previously reported looking into gender effects, we also compared four levels of internal regulation among the pupils and found that especially lower self-regulated pupils in “normal” science classes show a significantly higher self-regulated learning motivational behaviour in the outdoor educational setting. This may well be invoked by the physical activity inherent in the outdoor field research, a factor that is suggested to be specifically helpful for kids with attention deficits in the ordinary school context (Gapin & Etnier, 2010; Verret, Guay, Berthiaume, Gardiner, & Béliveau, 2012). Even if we have not tested the lower self-regulated pupils for any attention deficit disorder, and do not suppose highly external regulation as an indicator for any attention deficit syndrome, we could theorize that

similar activity treatment patterns do their work here. Further research with well-defined criteria concerning (the different types of) ADHD and external motivational behaviour would have to prove this hypothesis.

From an epistemological stance, it may be noteworthy that, as in “Alien at home...”, the use of NHST-techniques is justified by our critically small sample size, the descriptive analyses, the robust tests, and – again different from Fengler (2007) – the contextualization of the data with first-hand qualitative information providing logical support for the p-values and warranting also their practical significance.

HG₁: The pupils’ learning motivational behaviour measured in the context of the science lessons at school (NuT) and during the research week (FoWo) differs.	HG₂: The practical orientation of the programme at FoWo is perceived higher by the pupils than at NuT.	HG₃: Group dynamics and physical activity levels are relevant factors for positive motivational behaviour.
HS _{1.1} : The pupils’ learning motivational behaviour at FoWo is more self-regulated than at NuT.	HS _{2.1.1} : At FoWo, the pupils find that general rules can be better deferred from practical examples than at NuT.	HS _{3.1} : There is a positive correlation between self-regulated motivational behaviour and well-being within the group.
HS _{1.2} : Those pupils, who are less self-regulated at NuT profit especially from the outdoor teaching with respect to self-regulated learning.	HS _{2.1.2} : At FoWo, the pupils find the general applicability of the learned better than at NuT.	HS _{3.2} : There is a positive correlation between self-regulated motivational behaviour and enjoyment of the expedition.
	HS _{2.1.3} : At FoWo, in the perception of the pupils, the developing of new knowledge takes off more from own experiences than at NuT.	HS _{3.3} : There is a negative correlation between enjoyment of the expedition and the perceived exercise-level of the expedition.
	HS _{2.1.4} : At FoWo, in the perception of the pupils, the reference to every-day-life examples is higher than at NuT.	HS _{3.4} : There is a negative correlation between the perceived exercise-level of the expedition and self-regulated motivational behaviour.
	HS _{2.1.5} : At FoWo, the pupils find the specific applicability of the learned with respect to every-day-life-problems better than at NuT .	HS _{3.5} : There is a positive correlation between the perceived enjoyment of the expedition and self-regulated motivational behaviour.
	HS _{2.2.1} : At FoWo, there is a positive correlation between the pupils perception, that general rules can be deferred from practical examples, and highly self-regulated motivational behaviour.	
	HS _{2.2.2} : At FoWo, there is a positive correlation between the general applicability of the learned and highly self-regulated motivational behaviour.	
	HS _{2.2.3} : At FoWo, there is a positive correlation between the developing of new knowledge, which takes off more from own experiences, and highly self-regulated motivational behaviour.	
	HS _{2.2.4} : At FoWo, there is a positive correlation between the reference to every-day-life examples and highly self-regulated motivational behaviour.	
	HS _{2.2.5} : At FoWo, there is a positive correlation between the specific applicability of the learned with respect to every-day-life-problems and highly self-regulated motivational behaviour.	

Table 6. Hypotheses of “Investigating...”

4. Conclusion and Acknowledgements

This essay may be significantly longer than the required 25 pages. However, it was very important for me to present and explain the general context of my research in outdoor education, which is rather new and unknown in Germany, and to define my epistemological and philosophical stance. I believe this to be a quite common deficit in educational research, especially in outdoor education, which is dominated mainly by nature enthusiasts with little understanding for the complexity of “research in the mirror of nature.”

At the end of this essay, I need to say thank you to a lot of people and indicate their contributions to my doctorate dissertation.

First of all, many, many thanks need to be addressed to my parents, Helga and Herbert Dettweiler, who have not only supported me and my family financially, but who have also transcribed huge chunks of tape-recorded interview material and hand-written questionnaires. Without their support, this work would not have been possible.

My second thank-you goes to my wife Gabriele Lauterbach, who co-authored the two empirical papers. She collected data in the field by taking ethnographic notes and conducting interviews with the pupils for “Investigating...”. For both empirical papers, she has prepared the “classical” content analyses by coding the hand-written respectively transcribed material, and chose the relevant texts from the tremendous material. Furthermore, she translated those selected German texts into English for the sake of international dissemination of our research. More informally, Gabriele has been a very valuable critical and sceptical voice throughout the research projects.

Apart from her role as a scientist, Gabriele supported me emotionally with her love, her fabulous cooking, and her legendary sarcasm, especially after earlier versions of the two papers had been rejected in several review processes. Gabriele’s support from her parents, Werner and Christa Lauterbach, must not be forgotten here. Werner and Christa took turns with my parents taking care of our two daughters, Nauka and Smilla Ida, when both, Gabriele and I were in Berchtesgaden at the research weeks, or at international conferences to present the results of our research.

The third address of gratitude is professor Dr. Claudia Kugelmann, co-author in “Alien at home...” and “From the sea to the mountain...”, and mentor of my academic career. If not for Claudia, none of this had been written nor had there been any outdoor programming at TUM. She brought me back on track, so to speak, and connected my professional career in outdoor education to my (not so inexpensive) academic training after eight years in the private educational sector outside the ivory tower. Countless conversations and e-mails helped me to sharpen my academic profile and to emphasise her scientific heritage. Claudia’s contribution to “Alien at home...” is not only the financing and the steering of the team, but also the pointing at the directions of the research questions.

Those could not have been answered in such sharpness and precision if not professor Dr. Ali Ünlü had entered the team of my doctoral supervisors at a critical moment in November 2013, when I approached him with a set of questions regarding the “research week” data. Later in 2014, I included Ali also in the ERCS-project, after the first draft of our manuscript came back from review with the stipulation to accord the presentation of the data more according the U-curve theory. With his widely renowned mathematical competences and computing skills, Ali helped me to squeeze out the last bit of the data and “significantly” contributed to the internationally acknowledged quality of the two empirical papers.

Christoph Becker was one of my first students whom I supervised in their diploma theses. Christoph was of great help setting up the Forscherwochen in Berchtesgaden, and his academic background as a sports scientist qualified him to responsibly collect and prepare mainly the exercise physiological data for “Investigating...”.

Dr. Dr. Perikles Simon is an old friend of mine and professor of sports medicine at Mainz University. Peri’s contribution to “Alien at home...”, a topic rather odd for a world-class exercise physiologist and doping analyst, was twofold: firstly, he spent long hours with me discussing the epistemological and philosophical implications of empirical research – of facts and values, and of sense and non-sense of certain research approaches. Secondly, he introduced me into statistics (I had never had any formal training in empirical methodologies) and ran the first analyses of my KUS-ERCS data set with me stunningly looking over his shoulders. With his humour and experience with disinclined reviewers, Peri helped me to stay on the track Claudia had put me back on.

Dr. Jörg Wernecke, trained philosopher and member of the chair of Philosophy of Science at TUM, guided me as a second supervisor through the jungle of contemporary pragmatist thought. With remarkably few sessions over the course of two years, we managed to find a common way of understanding “understanding.”

Finally, Andrea Legl and Bernhard Gschrey helped with a few computations, tables and charts for the two empirical papers. Dr. Michael Streifinger contributed to the non-empirical essay “Expeditionary Learning...” with geographical excursion didactical text sequences.

I also want to thank the team at the Schülerforschungszentrum Berchtesgadener Land, and the teachers and pupils involved in the Forscherwochen for their motivation and courage to try out something new and unfamiliar.

Many conversations with Nils Faarlund, “norsk friluftsmann” and “vegleder” about the meaning of skiing, nature’s role in education, and the Norwegian mythology lead me to the general idea that manipulative abductive research methodology was most appropriate to apply in this field. As Askeladden, a(n anti) hero in many Norwegian folktales, we need to keep our eyes open for *patterns*, rather than blindly follow *rules* – which involves experience and intuition (Dreyfus, 1980). Meaning patterns derived

from such a “leug”⁹-epistemology is therefore most suitable knowledge to the educational researcher. Dr. Gunnar Breivik, professor emeritus at Norges Idrettshøgskole in Oslo, helped me to sharpen this rather narrative concept in philosophical terms.

Colleagues from Denmark (Dr. Peter Bentsen, Dr. Erik Mygind, and Dr. Reinhard Stelter) and Scotland (Dr. Simon Beames and Dr. Pete Allison) have contributed to this research informally with their thoughts in discussions and critical comments on early drafts of some of those texts constitutive of this thesis.

The planning of the research designs, choosing the methods of data collection and analysis, writing up the papers and serving as the corresponding author in the publication process – was my original work, together with the establishment and financing of a completely new approach to outdoor education, respectively, science teaching. And I sincerely hope that those final remarks signify the beginning of something new and lasting.

⁹ “Leug“: Used by skiers around Sondre Norheim in Telemark in the late 19th century to describe the coordinative competence needed when skiing down unknown steep terrain whilst intuitively reacting on the upcoming obstacles (Faarlund & Breivik, 2007).

5. Appendix

- I. Dettweiler, U., Kugelmann, C., & Streifinger, M. (2011). Expeditionary Learning - Unterwegs auf neuen pädagogisch-didaktischen Pfaden vom Meer bis in die Alpen [Expeditionary Learning - Underway on new pedagogical and didactical pathways from the sea into the Alps]. In H. D. Lange, G.; Leffler, T.; Siebe, A.; Zimlich, M. (Ed.), *Bewegungsbezogene Bildungskonzeptionen. Positionen und Analysen im Spannungsfeld zwischen Konzeption, Implementation und Evaluation. Jahrbuch Bewegungs- und Sportpädagogik in Theorie und Forschung der Deutschen Gesellschaft für Erziehungswissenschaft (DGfE)* (Vol. 10, pp. 129-143). Baltmannsweiler: Schneider.
- II. Dettweiler, U., Ünlü, A., Lauterbach, G., Legl, A., Simon, P., & Kugelmann, C. (2015). Alien at home: Adjustment strategies of students returning from a six-months over-sea's educational programme. *International Journal of Intercultural Relations*, 44(0), 72-87.
doi: <http://dx.doi.org/10.1016/j.ijintrel.2014.10.005>
- III. Dettweiler, U., Ünlü, A., Lauterbach, G., Becker, C., & Gschrey, B. (2015). Investigating the motivational behaviour of pupils during outdoor science teaching within self-determination theory. *Frontiers in Psychology*, 6.
doi: 10.3389/fpsyg.2015.00125

6. Bibliography

- Adler, N. J. (1981). Re-entry: Managing cross-cultural transitions. *Group & Organizational Studies*, 6(341-356).
- Adorno, T. W. (1976a). *The Positivist dispute in German sociology*. London: Heinemann.
- Adorno, T. W. (1976b). Sociology and Empirical Research (G. Adey & D. Frisby, Trans.). In G. Adey & D. Frisby (Eds.), *The Positivist Dispute in German Sociology* (pp. 68-86). London: Heinemann.
- Adorno, T. W. (2006). *Theorie der Halbbildung [Theory of half-education]*. Frankfurt am Main: Suhrkamp.
- Allison, P., Davis-Berman, J., & Berman, D. (2011). Changes in latitude, changes in attitude: analysis of the effects of reverse culture shock – a study of students returning from youth expeditions. *Leisure Studies*, 1–17. doi: 10.1080/02614367.2011.619011
- Allison, P., & Pomeroy, E. (2000). How shall we "know"?: Epistemological Concerns in Research in Experiential Education. *Journal of Experiential Education*, 23(2), 91–97.
- Apel, K.-O. (1998a). Can an ultimate foundation of knowledge be non-metaphysical? In K.-O. Apel & M. Papastephanou (Eds.), *From a transcendental-semiotic point of view* (pp. 81-102). Manchester, UK ; New York: Manchester University Press.
- Apel, K.-O. (1998b). The impact of analytical philosophy on my intellectual biography. In K.-O. Apel & M. Papastephanou (Eds.), *From a transcendental-semiotic point of view* (pp. 9–42). Manchester, UK ; New York: Manchester University Press.
- Apel, K.-O. (1998c). Transcendental semiotics and truth: the relevance of a Peircean consensus theory of truth in the present debate about truth theories. In K.-O. Apel & M. Papastephanou (Eds.), *From a transcendental-semiotic point of view* (pp. 64–80). Manchester, UK ; New York: Manchester University Press
- Apel, K.-O. (1998d). Wittgenstein and Heidegger: language forms and life forms. In K.-O. Apel & M. Papastephanou (Eds.), *From a transcendental-semiotic point of view* (pp. 122–159). Manchester, UK ; New York: Manchester University Press.

- Apel, K.-O., & Papastephanou, M. (Eds.). (1998). *From a transcendental-semiotic point of view*. Manchester, UK; New York: Manchester University Press.
- Asendorpf, J. B., & Aken, M. A. G. v. (1993). Deutsche Versionen der Selbstkonzeptskalen von Harter. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 25(1), 64-86.
- Atkinson, P. (1990). *The ethnographic imagination: Textual constructions of reality*. London: Routledge.
- Austin, J. L., & Warnock, G. J. (1962). *Sense and sensibilia*. Oxford: Clarendon Press.
- Ayer, A. J. (1961). *The foundations of empirical knowledge*. London: Macmillan.
- Bacon, S. (1987). *The Evolution of the Outward Bound Process*. Outward Bound USA Greenwich, Connecticut.
- Bateson, G. (1980). *Mind and nature: A necessary unity*. Great Britain: Wildwood House.
- Bauer, J. (2008). *Learning from Errors at Work. Studies on Nurses' Engagement in Error-Related Learning Activities*. (Doctoral Thesis), Universität Regensburg, Regensburg.
- Bauer, J., & Prenzel, M. (2012). European teacher training reforms. *Science*, 336, 1642-1643. doi: DOI:10.1126/science.1218387
- Bazeley, P. (2007). *Qualitative data analysis with NVivo*. Los Angeles ; London: SAGE.
- Becker, P. (2005). Das Abenteuer als eine Kategorie von Bildung [Adventure as a category of education]. In J. Bietz, R. Laging & M. Roscher (Eds.), *Bildungstheoretische Grundlagen der Bewegungs- und Sportpädagogik* (pp. 227-247). Baltmannsweiler: Schneider Verlag Hohengehren.
- Bedacht, A. (1994). *Erlebnispädagogik: Mode, Methode oder mehr? : Tagungsdokumentation des Forums Erlebnispädagogik* (2., unveränd. Aufl. ed.). München: Fachhochsch.-Schr. Sandmann.
- Bentsen, P., Mygind, E., & Randrup, T. B. (2009). Towards an understanding of udeskole: education outside the classroom in a Danish context. *Education*, 3(13 37 (1)), 29–44.
- Bentsen, P., Søndergaard Jensen, F., Mygind, E., & Barfoed Randrup, T. (2010). The extent and dissemination of udeskole in Danish schools. *Urban Forestry & Urban Greening*, 9(3), 235-243. doi: 10.1016/j.ufug.2010.02.001

- Bernstein, R. J. (2011). *Beyond Objectivism and Relativism: Science, Hermeneutics, and Praxis*: University of Pennsylvania Press, Incorporated.
- Biesta, G. J. J. (2010). Why 'What Works' Still Won't Work: From Evidence-Based Education to Value-Based Education. *Studies in Philosophy and Education*, 29(5), 491-503. doi: 10.1007/s11217-019-9191-x
- Biesta, G. J. J., & Burbules, N. C. (2003). *Pragmatism and educational research*. Oxford: Rowman & Littlefield Publishing Group, Incorporated.
- Bietz, J., Laging, R., & Roscher, M. (2005). *Bildungstheoretische Grundlagen der Bewegungs- und Sportpädagogik [Theoretical foundations of sports pedagogy and human kinetics]; [Expertentagung ... im Jahre 2002 in Marburg]*. Baltmannsweiler: Schneider Verl. Hohengehren.
- Bisson, C., & Luckner, J. (1996). Fun in Learning: The Pedagogical Role of Fun in Adventure Education. *Journal of Experiential Education*, 19(2), 108-112. doi: 10.1177/105382599601900208
- Bogner, A., Emerson, R. M., Fretz, R. I., & Shaw, L. L. (Eds.). (2002). *Das Experteninterview // Writing ethnographic fieldnotes: Theorie, Methode, Anwendung*. Opladen: Leske und Budrich University of Chicago Press.
- Böhm-Kasper, O., Schuchart, C., & Weishaupt, H. (2009). *Quantitative Methoden in der Erziehungswissenschaft*. Darmstadt: WBG.
- Böhme, G. (2002). *Die Natur vor uns. Naturphilosophie in pragmatischer Hinsicht [Nature in front of ourselves. Nature philosophy from a pragmatist point of view]*. Zug: Die Graue Edition.
- Böhme, G. (2003). *Leibsein als Aufgabe. Leibphilosophie in pragmatischer Hinsicht [Embodiment as human condition. Body-philosophy from a pragmatist point of view]*. Zug: Die Graue Edition.
- Bortz, J., Döring, N., & Bortz, D. (1995). *Forschungsmethoden und Evaluation : [für Sozialwissenschaftler] ; mit 60 Tabellen* (2., vollst. überarb. und aktualisierte Aufl. ed.). Berlin [u.a.]: Springer.
- Brabant, S., Palmer, C. E., & Gramling, R. (1990). Returning home: An empirical investigation of cross-cultural reentry. *International Journal of Intercultural Relations*, 14(4), 387-404. doi: [http://dx.doi.org/10.1016/0147-1767\(90\)90027-T](http://dx.doi.org/10.1016/0147-1767(90)90027-T)
- Brandt, A. (2005). *Förderung von Motivation und Interesse durch außerschulische Experimentierlabors [Enhancing motivation and interest through extra-school laboratories]*. Göttingen: Cuvillier.

- Brezing, H. (2000). Welche Bedürfnisse haben Anwender(innen), und wie werden sie in der Forschung abgedeckt? Die Bedeutung von Evaluationsstandards und von Effektivitätskriterien für die Praxis. In W. Hager, J.-L. Patry & H. Brezing (Eds.), *Evaluation psychologischer Interventionsmaßnahmen : Standards und Kriterien: ein Handbuch* (1. Aufl. ed., pp. 8-18). Bern [u.a.]: Huber.
- Brezinka, W. (1971). *Von der Pädagogik zur Erziehungswissenschaft eine Einführung in die Metatheorie der Erziehung*. Weinheim: Beltz.
- Bringer, J. D., Johnston, L. H., & Brackenridge, C. H. (2006). Using Computer-Assisted Qualitative Data Analysis Software to Develop a Grounded Theory Project. *Field Methods*, 18(3), 245-266. doi: 10.1177/1525822x06287602
- Brookes, A. (1991). *Thoughts on the conference theme*. Paper presented at the 7th National Outdoor Education Conference. Outdoor education, the quest for quality, Monash University, Frankston, Victoria, Australia.
- Brown, C. (1988). Internal Realism: Transcendental Idealism? *Midwest Studies in Philosophy*, 12, 145–155.
- Bühl, A., & Zöfel, P. (2005). *SPSS 12 Einführung in die moderne Datenanalyse unter Windows* (9., überarb. und erw. Aufl. ed.). München: Pearson Studium.
- Burnard, P. (1991). *Experiential learning in action*. Aldershot, Hants: Gower.
- Bushweller, K. (1997). Expeditions in learning. *The American School Board Journal*, 184(4), 30-33.
- Campbell, M., Cousins, E., Farren, G., Kamil, M., Larn, R., Rugen, L., & Udall, D. (1996). Expeditionary Learning Outward Bound Design. In S. Stringfield, S. Ross & L. Smith (Eds.), *Bold Plans for School Restructuring. The New American Schools Designs* (pp. 108-138). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Carver, R. (1996). Theory for practice: A framework for thinking about experiential education. *The Journal of Experiential Education*, 19(1), 8-13.
- Church, A. (1936). A note on the Entscheidungsproblem. *Journal of Symbolic Logic*, 1, 40–41.
- Cleveland, W. S., & Becker, A. (1996). *S-PLUS Trellis Graphics User's Manual. Versions 2.0 & 2.1*. MathSoft. Seattle, Washington.
- Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, 7, 249-253.

- Cohen, J. (1994). The Earth Is Round ($p < .05$) *American Psychologist*, 49(12), 997-1003.
- Cohen, J. (2010). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.
- Cohen, M. R., & Nagel, E. (1934). *An introduction to logic and scientific method*. New York: Harcourt.
- Cousins, E. (1999). Reflections on design principles. *Expeditionary Learning Outward Bound*, 7(5), 3-8.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.
- Cronbach, L. J., & Shavelson, R. J. (2004). My Current Thoughts on Coefficient Alpha and Successor Procedures. *Educational and Psychological Measurement*, 64(3), 391-418.
- Cronbach, L. J., & Suppes, P. (1969). *Research for tomorrow's schools: Disciplined inquiry in education*. New York: Macmillan.
- Davidson, D. (1984). *Inquiries into Truth and Interpretation*: Oxford University Press.
- Davidson, D., & Harman, G. (1970). Semantics of natural language. *Synthese*, 22(1-2), 1-2. doi: 10.1007/bf00413596
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum Publishing Co.
- Deci, E. L., & Ryan, R. M. (2000). The 'what' and 'why' of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11, 227-268.
- Deci, E. L., & Ryan, R. M. (2002). Overview of self-determination theory: An organismic dialectical perspective. In E. L. Deci, & Ryan, R. M. (Ed.), *Handbook of self-determination research* (pp. 3-33). Rochester: University of Rochester Press.
- Deci, E. L., & Vansteenkiste, M. (2004). Self-determination theory and basic need satisfaction: Understanding human development in positive psychology. *Ricerche di Psicologia*, 27, 17-34.
- Delaney, D., & Leitner, H. (1997). The political construction of scale. *Political Geography*, 16(2), 93-97.

- Dettweiler, U. (2012). The Journey Home. Some Thoughts on Psychological Adjustment Symptoms after Wilderness Expedition Programs. *Canadian Pathways Journal*, 24(2), 22-25.
- Dettweiler, U., & Kugelmann, C. (2010). The Pedagogical Value of Expeditions. A Little Rehearsal of Kurt Hahn's Concept of „Being Outward Bound“. *Outward Bound International Journal*, 24–28.
- Dettweiler, U., Kugelmann, C., & Streifinger, M. (2011). Expeditionary Learning - Unterwegs auf neuen pädagogisch-didaktischen Pfaden vom Meer bis in die Alpen [Expeditionary Learning - Underway on new pedagogical and didactical pathways from the sea into the Alps]. In H. D. Lange, G.; Leffler, T.; Siebe, A.; Zimlich, M. (Ed.), *Bewegungsbezogene Bildungskonzeptionen. Positionen und Analysen im Spannungsfeld zwischen Konzeption, Implementation und Evaluation. Jahrbuch Bewegungs- und Sportpädagogik in Theorie und Forschung der Deutschen Gesellschaft für Erziehungswissenschaft (DGfE)* (Vol. 10, pp. 129-143). Baltmannsweiler: Schneider.
- Dettweiler, U., Ünlü, A., Lauterbach, G., Becker, C., & Gschrey, B. (2015). Investigating the motivational behaviour of pupils during outdoor science teaching within self-determination theory. *Frontiers in Psychology*, 6. doi: 10.3389/fpsyg.2015.00125
- Dettweiler, U., Ünlü, A., Lauterbach, G., Legl, A., Simon, P., & Kugelmann, C. (2015). Alien at home: Adjustment strategies of students returning from a six-months over-sea's educational programme. *International Journal of Intercultural Relations*, 44(0), 72-87. doi: <http://dx.doi.org/10.1016/j.ijintrel.2014.10.005>
- Dettweiler, U., & Wolter, B. (2010). *Abenteuer Bildung - Ein pädagogisches Projekt unterwegs vom Meer in die Alpen*. Paper presented at the Bewegungsbezogene Bildungskonzeptionen. Zur Trias Konzeption, Implementation und Evaluation, Würzburg.
- Deusinger, I. M. (1986). *Frankfurter Selbstkonzept-Skalen*. Göttingen: Hogrefe.
- Deutscher Alpenverein. (2011). DAV-Unfallstatistik. München.
- Dews, P. (2001). Naturalismus und Anti-Naturalismus bei Habermas. *Deutsche Zeitschrift für Philosophie*, 49(6), 861-871.
- Dreier, W. (1960). Der Weg zur normativen Sozialwissenschaft. *Jahrbuch des Instituts für Christliche Sozialwissenschaft*, 1(1), 19-30.
- Dreyfus, H. L. (1980). Holism and Hermeneutics. *The Review of Metaphysics*, 34(1), 3-23.

- Düchting, F. (1994). Wenn Pädagogik was erleben will. Kritische Anmerkungen zum interessanten Verhältnis von "Erlebnispädagogik" und Jugendarbeit. *Deutsche Jugend*, 42(7-8), 318-326.
- Duerden, M. D., Widmer, M. A., Taniguchi, S. T., & McCoy, J. K. (2009). Adventures in Identity Development: The Impact of Adventure Recreation on Adolescent Identity Development. *Identity*, 9(4), 341-359. doi: 10.1080/15283480903422806
- Duncan, O. D. (1984). *Notes on social measurement : historical and critical*. New York: Russell Sage Foundation.
- Emerson, R. M. (2001). *Contemporary field research : perspectives and formulations* (2nd ed.). Prospect Heights, Ill.: Waveland Press.
- Engeln, K. (2004). *Schülerlabors: authentische, aktivierende Lernumgebungen als Möglichkeit, Interesse an Naturwissenschaften und Technik zu wecken [Student labs: authentic and activating learning environments as chance to enhance interest in science and technology]*. Berlin: Logos.
- Faarlund, N., & Breivik, G. (2007). Å være i bevegelse: Gunnar Breivik i samtale med Nils Faarlund. In A. M. Sookermany & J. W. Eriksen (Eds.), *Veglederen. Et festskrift til Nils Faarlund* (pp. 259–267). Oslo: Gan Aschehoug.
- Falco, E., & García-Lapresta, J. L. (2014). *Assigning numerical scores to qualitative scales*. Paper presented at the European Mathematical Psychology Group Meeting 2014, Tübingen.
- Fengler, J. (2007). *Erlebnispädagogik und Selbstkonzept. Eine Evaluationsstudie*. (Doctoral Dissertation), Logos-Verlag, Berlin.
- Fengler, J. (2012). Neue Forschung braucht das Land! Über die Vorzüge bivariater und multivariater Versuchspläne für die Erlebnispädagogik. *erleben und lernen. Internationale Zeitschrift für handlungsorientiertes Lernen*, 3&4, 26-29.
- Ferguson, A., Myers, C. S., Bartlett, R. J., Banister, H., Bartlett, F. C., Brown, W., . . . Tucker, W. S. (1940). Quantitative Estimates of Sensory Events. *Advancement of Science*, 2, 331-349.
- Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh: Oliver & Boyd.
- Flick, U. (2007). *Qualitative Sozialforschung: Eine Einführung [Qualitative Social Research: An Introduction]* (Orig.-Ausg., vollst. überarb. und erw. Neuausg. ed.). Reinbek bei Hamburg: Rowohlt-Taschenbuch-Verl.

- Flick, U., Kardorff, E. v., & Steinke, I. (Eds.). (2004). *A companion to qualitative research*. Estados Unidos: Sage Publications.
- Gadamer, H.-G. (1965). *Wahrheit und Methode Grundzüge einer philosophischen Hermeneutik* (2., durch einen Nachtrag erw. Aufl. ed.). Tübingen: Mohr & Siebeck.
- Galuske, M. (2003). *Methoden der sozialen Arbeit : eine Einführung* (5. Aufl ed.). Weinheim [u.a.]: Juventa-Verl.
- Gapin, J., & Etnier, J. L. (2010). The Relationship Between Physical Activity and Executive Function Performance in Children With Attention-Deficit Hyperactivity Disorder. *Journal of Sport & Exercise Psychology*, 43(6), 753-763.
- Garson, D. (2012). *Testing Statistical Assumptions* Retrieved from <http://www.statisticalassociates.com/assumptions.pdf>
- Garz, D., & Kraimer, K. (Eds.). (1994). *Die Welt als Text: Theorie, Kritik und Praxis der objektiven Hermeneutik* (1 ed.). Frankfurt am Main: Suhrkamp.
- Gaw, K. F. (2000). Reverse culture shock in students returning from overseas. *International Journal of Intercultural Relations*, 24(1), 83-104. doi: [http://dx.doi.org/10.1016/S0147-1767\(99\)00024-3](http://dx.doi.org/10.1016/S0147-1767(99)00024-3)
- Gibbs, G. (2002). *Qualitative data analysis : explorations with NVivo*. Buckingham: Open University.
- Gillis, L., & Speelman, E. (2008). Are Challenge (Ropes) Courses an Effective Tool? A Meta-Analysis. *Journal of Experiential Education*, 31(2), 111-135. doi: 10.5193/JEE.31.2.111
- Gilsdorf, R. (1999). *Abenteuer Schule* (1. Aufl ed.). Alling: Sandmann.
- Glock, H.-J. (1996). *A Wittgenstein dictionary*. Oxford: Blackwell Reference.
- Gookin, J. (2011). *Development of a psychometric scale to measure challenge (stress) type and intensity in wilderness education students*. (Doctoral Dissertation), Prescott College, Tucson, Arizona. Retrieved from <http://gradworks.umi.com/3465932.pdf> (3465932)
- Guba, E. G., & Lincoln, Y. S. (1982). Epistemological and methodological bases of naturalistic inquiry. *ECTJ*, 30(4), 233-252. doi: 10.1007/BF02765185
- Guderian, P. (2007). *Wirksamkeitsanalyse außerschulischer Lernorte - der Einfluss mehrmaliger Besuche eines Schülerlabors auf die Entwicklung des Interesses an Physik [How effective are non-formal learning environments - the*

influence of multiple visits in a student research lab to the development of interest in physics]. Humboldt-Universität zu Berlin.

- Habermas, J. (1984). *Vorstudien und Ergänzungen zur Theorie des kommunikativen Handelns*. Frankfurt am Main: Suhrkamp.
- Habermas, J. (1992). Bemerkungen zu J. Searles "Meaning, Communication, and Representation". In J. Habermas (Ed.), *Nachmetaphysisches Denken* (pp. 136-149). Frankfurt a.M.: Suhrkamp.
- Habermas, J. (1995). Wahrheitstheorien. In J. Habermas (Ed.), *Vorstudien und Ergänzungen zur Theorie des kommunikativen Handelns* (1 ed., pp. 127–183). Frankfurt am Main: Suhrkamp.
- Habermas, J. (1999a). *Theorie des kommunikativen Handelns* (3 ed.). Frankfurt am Main: Suhrkamp Taschenbuch Verlag.
- Habermas, J. (1999b). *Wahrheit und Rechtfertigung. Philosophische Aufsätze*. Frankfurt am Main: Suhrkamp.
- Habermas, J. (2001). *On the pragmatics of social interaction : preliminary studies in the theory of communicative action*. Cambridge, Mass.: MIT Press.
- Hammersley, M., & Atkinson, P. (2007). *Ethnography: Principles in practice* (3 ed.). London ;, New York: Routledge.
- Heckmair, B., & Michl, W. (2012 (1993)). *Erleben und Lernen : Einführung in die Erlebnispädagogik* (7. Aufl. ed.). München: Reinhardt.
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24, 417-441, and 498-520.
- Hufenus, H.-P. (1997). Erlebnispädagogik - Grundlagen. In F. Herzog (Ed.), *Erlebnispädagogik: Schlagwort oder Konzept?* (pp. 85-99). Luzern: Edition der Schweizerischen Zentralstelle für Heilpädagogik (SZH).
- Hufenus, H.-P. (2009). *Handbuch für Outdoor-Guides : Theorie und Praxis der Outdoorleitung* (3., überarb. Aufl. ed.). Augsburg: ZIEL.
- James, T. (1980). *Can the Mountains Speak for Themselves?* Colorado Outward Bound School. Denver. Retrieved from <http://www.wilderdom.com/facilitation/Mountains.html>
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher*, 33(7), 14-26. doi: 10.3102/0013189x033007014

- Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a Definition of Mixed Methods Research. *Journal of Mixed Methods Research, 1*(2), 112-133. doi: 10.1177/1558689806298224
- Jones, K. (1998). Scale as epistemology. *Political Geography, 17*(1), 25–28.
- Kant, I. (1957). *Prolegomena zu einer jeden künftigen Metaphysik, die als Wissenschaft wird auftreten können (1783)* (reprint ed.). Hamburg: Meiner.
- Kaplan, S., & Talbot, J. F. (1983). Psychological benefits of a wilderness experience. In I. Altman & J. F. Wohlwill (Eds.), *Human behaviour and environment: Advances in theory and research* (Vol. 163-205). New York: Plenum Press.
- Kasper, D., & Ünlü, A. (2013). On the relevance of assumptions associated with classical factor analytic approaches. *Frontiers in Psychology, 4*(Article 109), 1-20. doi: 10.3389/fpsyg.2013.00109
- Kelly, F. J. (1974). *Outward Bound and Delinquents: A ten year experience*. Paper presented at the Conference on Experiential Education, Estes Park, CO.
- Kelly, F. J., & Baer, D. J. (1968). *Outward Bound schools as an alternative to institutionalization for adolescent delinquent boys*. Boston: Fandel Press.
- Kelvin, W. T. (1891). *Popular lectures and addresses* (2. ed.). London: Macmillan and Co.
- Klafki, W. (2005). Bewegungskompetenz als Bildungsdimension [Movement Literacy as a Dimension of Education]. In R. Laging & R. Prohl (Eds.), *Bewegungskompetenz als Bildungsdimension. Reprint ausgewählter Beiträge aus den dvs-Bänden 104 und 120* (pp. 15-24). Hamburg: Czwalina Verlag.
- Klafki, W. (2007). *Neue Studien zur Bildungstheorie und Didaktik: Zeitgemässe Allgemeinbildung und kritisch-konstruktive Didaktik [New studies on the theory of education and didactics: Contemporary literacy and critical-constructive didactics]* (6 ed.). Weinheim: Beltz.
- Klafki, W., & Braun, K.-H. (2007). *Wege pädagogischen Denkens. Ein autobiografischer und erziehungswissenschaftlicher Dialog [Ways of pedagogical thought. An autobiographical and educational dialogue]*. München: Reinhardt.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3. ed.). New York: Guilford Press.
- Koepping, K. P. (1987). Authentizität als Selbstfindung durch den anderen. Ethnologie zwischen Engagement und Reflexion, zwischen Leben und Wissenschaft [Authenticity as self-perception through the other. Ethnology between engagement and reflexion, between life and science]. In H. P. Duerr

- (Ed.), *Authentizität und Betrug in der Ethnologie* (1 ed., pp. 7–37). Frankfurt am Main: Suhrkamp.
- Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260), 583-621.
- Kugelmann, C., & Lauterbach, G. (2011). KUS-Projekt: Abschlussbericht an die Staedtler-Stiftung. München; Erlangen: Technische Universität München/Friedrich-Alexander-Universität Erlangen/Nürnberg.
- Leberman, S. I., & Martin, A. J. (2003). Does pushing comfort zones produce peak learning experiences? *Australian Journal of Outdoor Education*, 7(1), 10-19.
- Lefebvre, H. (1991). *The production of space* (D. Nicholson-Smith, Trans.). Oxford: Blackwell.
- Lenk, H. (1993). *Interpretationskonstrukte zur Kritik der interpretatorischen Vernunft*. Frankfurt am Main: Suhrkamp.
- Lenk, H., & Maring, M. (1997). Welt ist real, aber Welterfassung interpretativ. Zur Reichweite der interpretatorischen Erkenntnis. In B. Frierberthäuser & A. Prengel (Eds.), *Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft* (pp. 209-220). Weinheim, München: Juventa Verlag.
- Levesque, C., Zuehlke, A. N., Stanek, L. R. & Ryan, R. M. . (2004). Autonomy and competence in German and American university students: A comparative study based on the self-determination-theory. *Journal of Educational Psychology*, 96(1), 68-84.
- Levin, S. A. (1992). The problem of pattern and scale in ecology: the Robert H. MacArthur Award Lecture. *Ecology*, 73(6), 943–1967.
- Lewis, R. B. (2004). NVivo 2.0 and ATLAS.ti 5.0: A Comparative Review of Two Popular Qualitative Data-Analysis Programs. *Field Methods*, 16(4), 439-464. doi: 10.1177/1525822x04269174
- Liebau, E. (1987). *Gesellschaftliches Subjekt und Erziehung. Zur pädagogischen Bedeutung der Sozi-alisationstheorien von Pierre Bourdieu und Ulrich Oevermann*. Weinheim; München: Juventa-Verlag.
- Liessmann, K. P. (2006). *Theorie der Unbildung: Die Irrtümer der Wissensgesellschaft [Theory of ignorance. The errors of the information society]*. Wien: Zsolnay.
- Luce, D. R. (1995). Four tensions concerning mathematical modeling in psychology. *Annual Review of Psychology*, 46, 1–26. doi: 10.1146/annurev.ps.46.020195.000245

- Magnani, L. (2001). *Abduction, reason, and science. Processes of discovery and explanation*. New York: Kluwer.
- Maiwald, K.-O. (2005). Competence and Praxis: Sequential Analysis in German Sociology. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 6(3), Art. 31.
- Mann, H. B., & Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Annals of Mathematical Statistics*, 18(1), 50-60. doi: 10.1214/aoms/1177730491
- Manson, S. M. (2008). Does scale exist? An epistemological scale continuum for complex human–environment systems. *Geoforum*, 39(2), 776-788. doi: 10.1016/j.geoforum.2006.09.010
- Marceau, D. J., & Hay, G. J. (1999). Remote sensing contributions to the scale issue. *Canadian Journal of Remote Sensing*, 25(4), 357–366.
- Marston, S. A. (2000). The social construction of scale. *Progress in Human Geography*, 24(2), 219-242.
- Martin, A. J. (2001). *Towards the next generation of experiential education programmes: A case study of Outward Bound*. (Doctoral Thesis), Massey University, Palmerston North, New Zealand.
- Martin, J. N. (1984). The intercultural reentry: Conceptualization and directions for future research. *International Journal of Intercultural Relations*, 8(2), 115-134. doi: [http://dx.doi.org/10.1016/0147-1767\(84\)90035-X](http://dx.doi.org/10.1016/0147-1767(84)90035-X)
- Mayer, V. (1997). *Semantischer Holismus. Eine Einführung*. Berlin: Akademie-Verlag.
- Mayring, P. (2014). *Qualitative content analysis: theoretical foundation, basic procedures and software solution*. Klagenfurt: Beltz.
- Meehl, P. E. (1990a). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry*, 1, 108-141.
- Meehl, P. E. (1990b). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports (Monograph Suppl.)*, 66(I-V66), 195-244.
- Messer, J. A. (2007). *Das Realismus-Problem der Quantenmechanik angesichts der Dekohärenz-Interpretation*. (Doctoral Thesis), Justus-Liebig-Universität Gießen, Gießen. Retrieved from <http://geb.uni-giessen.de/geb/volltexte/2008/6300/pdf/MesserJoachim-2008-03-26.pdf>

- Mjøset, L. (2000). Hvorfor er sosiologiske teoretikere så uenige? *Appollon Forskningsmagasinet*, (2). Retrieved from Appolon website: <http://www.apollon.uio.no/artikler/2000/sosiologer.html>
- Müller, F., Hanfstingl, B., & Andreitz, I. (2007). Skalen zur motivationalen Regulation beim Lernen von Schülerinnen und Schülern. Adaptierte und ergänzte Version des Academic Self-Regulation Questionnaire (SRQ-A) nach Ryan & Connell [Scales of motivational learning regulation of pupils. Adapted and extended version of the Academic Self-Regulation Questionnaire (SRQ-A) of Ryan & Connell]. In Institut für Unterrichts- und Schulentwicklung (Ed.), *Wissenschaftliche Beiträge*. Klagenfurt: Alpen-Adria-Universität.
- Mygind, E. (2007). A comparison between children's physical activity levels at school and learning in an outdoor environment. *Journal of Adventure Education and Outdoor Learning*, 2(7), 161-176. doi: 10.1080/14729670701717580
- Neill, J. T. (2002/2008, 1.11.2011). Meta-Analytic Research on the Outcomes of Outdoor Education: Paper presented to the 6th Biennial Coalition for Education in the Outdoors Research Symposium, Bradford Woods, IN, 11-13 January, 2002, updated 2008. from http://www.outward-bound.org/staff_sub2_research.htm
- Neill, J. T. (2004). Outdoor Education: Similar and Related Terms. Retrieved September 11, 2012, from <http://wilderdom.com/definitions/definitionstable.html>
- Neill, J. T. (2006). History of Outdoor Education Research. Retrieved September 11, 2012, from <http://wilderdom.com/research/HistoryResearch.html>
- Neyman, J., & Pearson, E. S. (1933). On the Problem of the Most Efficient Tests of Statistical Hypotheses. *Philosophical Transactions of the Royal Society*, 231, 289-337. doi: 10.1098/rsta.1933.0009
- Nicol, R. (2002a). Outdoor education: Research topic or universal value? Part one. *Journal of Adventure Education & Outdoor Learning*, 2(1), 29-41. doi: 10.1080/14729670285200141
- Nicol, R. (2002b). Outdoor education: Research topic or universal value? Part Two. *Journal of Adventure Education & Outdoor Learning*, 2(2), 85-99. doi: 10.1080/14729670285200201
- Nicol, R. (2003). Outdoor education: Research topic or universal value? Part three. *Journal of Adventure Education & Outdoor Learning*, 3(1), 11-27. doi: 10.1080/14729670385200211

- Oberg, K. (2006(1960)). Cultural Shock: Adjustment to New Cultural Environments. Reprint from *Practical Anthropology* 7 (177-182). *curare*, 29(2+3), 142-146.
- Odden, A. (2001). *Vitenskapsteoretiske posisjoner innen norsk friluftslivsforskning*. Høgskolen i Telemark, Bø. Retrieved from <http://hdl.handle.net/2282/712>
- Oevermann, U. (1979a). Die Methodologie einer "objektiven Hermeneutik" und ihre allgemeine for-schungslogische Bedeutung in den Sozialwissenschaften. In H.-G. Soeffner (Ed.), *Interpretative Verfahren in den Sozial- und Textwissenschaften* (pp. 352–434). Stuttgart: Metzler.
- Oevermann, U. (1979b). Die Methodologie einer "objektiven Hermeneutik" und ihre allgemeine forschungslogische Bedeutung in den Sozialwissenschaften. In H.-G. Soeffner (Ed.), *Interpretative Verfahren in den Sozial- und Textwissenschaften* (pp. 352–434). Stuttgart: Metzler.
- Oevermann, U. (1996, 2002). Klinische Soziologie auf der Basis der Methodologie der objektiven Hermeneutik – Manifest der objektiv hermeneutischen Sozialforschung. from http://www.ihs.de/publikationen/Ulrich_Oevermann-Manifest_der_objektiv_hermeneutischen_Sozialforschung.pdf
- Park, H. S., Dailey, R., & Lemus, D. (2002). The use of exploratory factor analysis and principal components analysis in communication research. *Human Communication Research*, 28(4), 562-577.
- Passarelli, A., Hall, E., & Anderson, M. (2010). A Strengths-Based Approach to Outdoor and Adventure Education: Possibilities for Personal Growth. *Journal of Experiential Education*, 33(2), 120-135. doi: 10.5193/jee33.2.120
- Pearl, J. (2009). *Causality : models, reasoning, and inference* (2nd ed.). Cambridge: Cambridge University Press.
- Pearson, K. (1900). χ . On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series 5*, 50(302), 157-175. doi: 10.1080/14786440009463897
- Peirce, C. S. (1955). Abduction and Induction. In J. Buchler (Ed.), *Philosophical Writings of Peirce* (pp. 5-22). New York: Dover Publications.
- Plackett, R. L. (1983). Karl Pearson and the Chi-Squared Test. *International Statistical Review / Revue Internationale de Statistique*, 51(1), 59-72.
- Pollard, P., & Richardson, J. T. E. (1987). On the probability of making Type I errors. *Psychological Bulletin*, 102, 159-163.

- Porst, R. (2001). *Wie man die Rücklaufquote bei postalischen Befragungen erhöht [How to increase the return quote in postal survey]* (Vol. 9). Mannheim: Zentrum für Umfragen, Methoden und Analysen.
- Praxisfeld. (2002). *Informationsmaterial*. Praxisfeld GmbH. Hamburg.
- Praxisfeld. (2006). *Drum prüfe, wer ans Seil sich bindet. Einführung in die Arbeit mit stationären Ropes Courses*. Augsburg: Ziel-Verlag.
- Prenzel, M. (2009). Challenges facing the educational system. In European Science Foundation (Ed.), *Vital questions: The contribution of European social science* (pp. 30-33). Strasbourg.
- Priest, S. (1987). *Preparing effective outdoor pursuit leaders*. Eugene, Oregon: Institute of Recreation Research and Service, Department of Leisure Studies and Services, University of Oregon.
- Priest, S., & Gass, M. A. (2005 (1997)). *Effective leadership in adventure programming* (2nd ed.). Champaign, IL: Human Kinetics.
- Putnam, H. (1981). *Reason, truth and history*. Cambridge: Cambridge University Press.
- Putnam, H. (1983). *Realism and reason*. Cambridge: Cambridge University Press.
- Putnam, H. (1987). *The many faces of realism : the Paul Carus lectures*. La Salle, Ill.: Open Court.
- Putnam, H. (1988). *Representation and reality*. Cambridge, Mass ; London: MIT Press.
- Putnam, H., & Conant, J. (1990). *Realism with a human face*. Cambridge, Mass ; London: Harvard University Press.
- Rakoczy, K., Buff, A., & Lipowsky, F. (2005). *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie "Unterrichtsqualität, Lernverhalten und mathematisches Verständnis" [Documentation of the inventories and data interpretation schemes of the Swiss-German video-study "Teaching Quality, Learning Behaviour, and Mathematical Literacy"]*. Teil 1. *Befragungsinstrumente. Materialien zur Bildungsforschung (13)*. Frankfurt am Main: Gesellschaft zur Förderung Pädagogischer Forschung / Deutsches Institut für Internationale Pädagogische Forschung.
- Rakoczy, K., Klieme, E., & Pauli, C. (2008). Die Bedeutung der wahrgenommenen Unterstützung motivationsrelevanter Bedürfnisse und des Alltagsbezugs im Mathematikunterricht für die selbstbestimmte Motivation [The meaning of perceived support for motivational needs and the practice orientation in maths

- teaching for self-determined motivation]. *Zeitschrift für Pädagogische Psychologie*, 22(1), 25-35.
- Reichertz, J. (1994). Von Gipfeln und Tälern. Bemerkungen zu einigen Gefahren, die den objektiven Hermeneuten erwarten. In D. Garz & K. Kraimer (Eds.), *Die Welt als Text* (1 ed., pp. 125–152). Frankfurt am Main: Suhrkamp.
- Reiners, A., & Schmieder, W. (1997). *Neue Sammlung motivierender Interaktionsspiele* (4. Aufl ed.). Alling: Sandmann.
- Researchware. (1998). *HyperRESEARCH*. Researchware. Randolph, MA.
- Richards, T. (2004). History. Retrieved October 9, 2012, from <http://www.outwardbound.net/about/history.html>
- Rodgers, J. L. (2010). The epistemology of mathematical and statistical modeling: A quiet methodological revolution. *American Psychologist*, 65(1), 1-12. doi: 10.1037/a0018326
- Rohrlich, B. F., & Martin, J. N. (1991). Host country and reentry adjustment of student sojourners. *International Journal of Intercultural Relations*, 15(2), 163-182. doi: [http://dx.doi.org/10.1016/0147-1767\(91\)90027-E](http://dx.doi.org/10.1016/0147-1767(91)90027-E)
- Rorty, R. (1980). *Philosophy and the mirror of nature* (2 ed.). Princeton, N.J: Princeton University Press.
- Rorty, R. (1989). *Contingency, irony, and solidarity* (3 ed.). Cambridge: Cambridge Univ. Press.
- Rorty, R. (1991a). Inquiry as recontextualization: An anti-dualist account of interpretation. In R. Rorty (Ed.), *Objectivity, relativism, and truth* (pp. 93–110). Cambridge: Cambridge University Press.
- Rorty, R. (1991b). Pragmatism, Davidson, and Truth. In R. Rorty (Ed.), *Objectivity, relativism, and truth* (pp. 126–150). Cambridge: Cambridge University Press.
- Rössler, P. (2010). *Inhaltsanalyse* (2 ed.). Konstanz: UVK-Verlagsgesellschaft.
- Ryan, R. M., & Connell, J. P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology*, 57(5), 749-761. doi: 10.1037/0022-3514.57.5.749
- Saint-Mont, U. (2011). *Statistik im Forschungsprozess. Eine Philosophie der Statistik als Baustein einer integrativen Wissenschaftstheorie*. Heidelberg: Physica-Verlag.
- Sarkar, D. (2008). *Lattice. Multivariate data visualization with R*. New York, NY: Springer.

- Satorra, A. (2000). Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In R. D. H. Heijmans, D. S. G. Pollock & A. Satorra (Eds.), *Innovations in multivariate statistical analysis. A Festschrift for Heinz Neudecker* (pp. 233-247). London: Kluwer Academic Publishers.
- Satorra, A., & Bentler, P. M. (2010). Ensuring Positiveness of the Scaled Difference Chi-square Test Statistic. *Psychometrika*, 75(2), 243-248.
- Schaad, N. (1993). Erleben und miteinander reden. Reflexionsmodelle in der Erlebnispädagogik. *e&l. Zeitschrift für handlungsorientierte Pädagogik*, 2 and 3, 49-53.
- Schödlbauer, C. (1999). *Metaphern Schnellstrassen, Saumpfade und Sackgassen des Lernens*. Augsburg: ZIEL.
- Schödlbauer, C. (2000). *Metaphorisches Lernen in erlebnispädagogischen Szenarien eine Untersuchung über handlungsorientierte Lehr-Lern-Prozesse*. (Diss Univ Erlangen-Nürnberg, 2000), Kova*c, Hamburg.
- Searle, J. R. (1969). *Speech acts : an essay in the philosophy of language*. Cambridge: Cambridge University Press.
- Searle, J. R. (1999). *Mind, language and society : doing philosophy in the real world*. London: Weidenfeld & Nicolson.
- Silverman, D. (2011). *Interpreting qualitative data : a guide to the principles of qualitative research* (4th ed.). Los Angeles ; London: SAGE.
- Skjervheim, H. (1982). Structuralism, empiricism and intentionalism. In I. Gullvåg & J. Wetlesen (Eds.), *In sceptical wonder* (pp. 129–146). Oslo ;, Bergen ;, Tromsø: Universitetsforlaget.
- Sproule, J., Martindale, R., Wang, J., Allison, P., Nash, C., & Gray, S. (2013). Investigating the experience of outdoor and adventurous project work in an educational setting using a self-determination framework. *European Physical Education Review*, 19(3), 315-328. doi: 10.1177/1356336X13495629
- Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science*, 103, 677-680.
- Streifinger, M. W. (2010). *Praxisbeispiel einer geodidaktischen Exkursion zur Optimierung des glazialmorphologischen Verständnisses im Untersuchungsgebiet Hoher Kranzberg, Mittenwald, Wallgau empirische Untersuchung zur Exkursionsdidaktik*.
- Tamhane, A. C. (1979). A Comparison of Procedures for Multiple Comparisons of Means with Unequal Variances. *Journal of the American Statistical Association*, 74(366), 471-480. doi: 10.2307/2286358

- Thomas, A. E., & Müller, F. H. (2014). Autonomy support: a key for understanding students learning motivation in science? *Zeitschrift für Bildungsforschung*, 4(1), 43-61. doi: 10.1007/s35834-013-0073-5
- Topitsch, E., & Payer, P. (1993). *Logik der Sozialwissenschaften* (12 ed.). Königstein/Ts.: Athenäum.
- Tukey, J. W. (1986 (1961)). Statistical and Quantitative Methodology. In L. V. Jones (Ed.), *The collected works of J. W. Tukey* (Vol. III: Philosophy and Principles of Data Analysis: 1949-1964., pp. 143-181). London: Chapman & Hall.
- Turing, A. (1937). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, 2(42), 230–265.
- Van der Smissen, B., & Brookhiser, J. (1982 (1972)). *Bibliography of research : organized camping, environmental education, adventure activities, interpretative services, outdoor recreation users & programming*. Martinsville, IN (Bradford Woods, Martinsville 46151): American Camping Association.
- Verret, C., Guay, M.-C., Berthiaume, C., Gardiner, P., & Béliveau, L. (2012). A Physical Activity Program Improves Behavior and Cognitive Functions in Children With ADHD: An Exploratory Study. *Journal of Attention Disorders*, 16(1), 71-80. doi: 10.1177/10870547110379735
- Vollmer, G. (1994). *Evolutionäre Erkenntnistheorie. Angeborene Erkenntnisstrukturen im Kontext von Biologie, Psychologie, Linguistik, Philosophie und Wissenschaftstheorie* (6 ed.). Stuttgart: Hirzel.
- Vossenkuhl, W. (1998). ‚Verstehen‘ verstehen. Über Analyse und Hermeneutik. In B. Kanitscheider, F. J. Wetz & B. Suchan (Eds.), *Hermeneutik und Naturalismus* (pp. 168–191). Tübingen: Mohr Siebeck.
- Warner, A. (1984). How to creatively evaluate programs. *Journal of Experiential Education*, 7(2), 38-43.
- Weber, M., & Winkelmann, J. (1951). *Gesammelte Aufsätze zur Wissenschaftslehre* (2. erw. u. verb. Aufl. ed.). Tübingen: J. C. B. Mohr (Paul Siebeck).
- Wernecke, J. (2007). *Handeln und Bedeutung: L. Wittgenstein, Ch. S. Peirce und M. Heidegger zu einer Propädeutik einer hermeneutischen Pragmatik*. Berlin: Duncker & Humblot.
- White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review*, 66, 297–333.

- Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6), 80-83.
- Wirth, U. (2003). Die Phantasie des Neuen als Abduktion. *Deutsche Vierteljahrsschrift für Literatur und Geistesgeschichte*, 77(4), 591–618.
- Wirth, U. (2008). Die Konjektur als blinder Fleck einer Geschichte bedingten Wissens. In C. Welsh & S. Willer (Eds.), *"Interesse für bedingtes Wissen"* (pp. 269–294). München: Fink.
- Wittgenstein, L. (1967). *Philosophical investigations* (2nd ed.). Oxford: Blackwell.
- Wittgenstein, L. (1984a). Tractatus logico-philosophicus. In J. Schulte (Ed.), *Ludwig Wittgenstein. Werkausgabe* (1 ed., Vol. 1). Frankfurt am Main: Suhrkamp.
- Wittgenstein, L. (1984b). Vermischte Bemerkungen. In J. Schulte (Ed.), *Ludwig Wittgenstein. Werkausgabe* (1 ed., Vol. 8). Frankfurt am Main: Suhrkamp.
- Wolcott, H. F. (1994). *Transforming qualitative data : description, analysis, and interpretation*. Thousand Oaks, Calif.: Sage Publications.
- Ziegenspeck, J. (1992). *Erlebnispädagogik : Rückblick - Bestandsaufnahme - Ausblick ; Bericht über den gegenwärtigen Entwicklungsstand der Erlebnispädagogik in der Bundesrepublik Deutschland unter besonderer Berücksichtigung der Lüneburger Anstöße und Projekte ; Dokumentation der geleisteten praktischen und theoretischen Arbeit (1980 - 1992)* (4. Aufl ed.). Lüneburg: Ed. Erlebnispädagogik.