A Systematic Approach on the Theoretical Quality of Educational Intervention Research: The Intervention Theory Questions (ITQ)

Hermann Astleitner
University of Salzburg

Paper presented at
European Conference on Educational Research (ECER)
11-15 September 2013, Istanbul (Turkey)

Abstract

Intervention research is an important field in establishing quality assurance procedures and related professional principles in educational applied research and practice. Recently, conceptualizing and developing educational intervention research guidelines and activities were, above all, in the hands of social research methodologists handling different criteria of empirical research, for example, validity or design sensitivity together with related assessment tools. However, improvements in educational intervention research also could come from considering expanded theoretical perspectives. It is the purpose of this paper to develop and apply a tool (called Intervention Theory Questions (ITQ)) that should assist educational researchers and evaluators in assessing the theoretical quality of educational interventions. The tool was based on systematically scanning, evaluating, and integrating literature on theory development and intervention design questions. It consists of more than 50 questions related to problem solution (exploration, goal building, triangulation, progress, theory evaluation), theory development (covariation, other causes eliminated, process explanation, contrasting conditions), intervention design (active and passive ingredients, selection of participants, possible modalities, bias handling), and discussions (effectiveness, side effects, follow ups, other contexts). The application of the tool is demonstrated by showing how the design and evaluation of an educational intervention in the field of aggressive behavior prevention changes when considering ITQ. Finally, open research questions are briefly discussed.

Keywords: educational effectiveness, assessment tools, experimental design, theory development,

A Systematic Approach on the Theoretical Quality of Educational Intervention Research: The Intervention Theory Questions (ITQ)

Repeatedly, scientists, decision makers, and professional organizations in the field of educational evaluation have stressed the need for more and better research on the effectiveness of educational interventions (e.g., Kelly, Baek, Lesh, & Bannan-Ritland, 2008). Within such interventions, it is tested resp. evaluated whether a certain activity, for example, an instruction, a training, or a therapy can effectively increase or decrease selected goal behaviors in educational settings (e.g., daily instruction in schools, continuing education activities, or counseling programs). Knowledge about the effectiveness of interventions helps in improving research, but also in establishing evidence-based decision making as a long-term objective in educational practice. Ways of evidence-based decision making are well established in medical diagnosis, crime risk assessment, credit-rating of companies, industrial process operation, and so on, but still remain to be significantly implemented in educational settings (e.g., Pourret, Naim, & Marcot, 2008).

In order to optimize educational intervention research, researchers are considering methodological standards of experimental or quasi-experimental research, most of them related to traditional work from Campbell (1957) or Weber and Cook (1972). These standards concern internal validity, external validity, and design sensitivity for research designs, or treatment validity, treatment integrity, and strength of treatment for treatment design (e.g., Fuchs, Fuchs, & Speece, 2002; Hagermoser Sanetti & Kratochwill, 2009; Highhouse, 2009; Lipsey, 1990; Mitchell & Jolley, 2010; Yeaton & Sechrest, 1981). Valentine and Cooper (2008) have integrated many of these standards and delivered a comprehensive "assessment device" for evaluating the methodological quality of intervention research activities. Having such standards and practicable assessment tools and using them when planning, conducting, and evaluating educational interventions should improve activities and effectiveness considerably.

However, there is also another source for optimizing intervention research. This other source comes from considering theories. Educational theories are nomological theories (if-then-statements in, for example, a theory of learning), but also technological theories (if-do-statements in, for example, an instructional design theory). Although these types of theories differ in respect to goals (e.g., finding truth in basic research versus being effective in applied research), both are often mixed up in daily educational research (Reigeluth, 1999, p. 12; see also for a comprehensive discussion: Patry & Perrez, 2000). Undoubtedly, theories help to optimize educational interventions and are therefore crucial for educational decision-making, because they assist, for example, in developing innovations, discovering problems, in finding causes for problems, in designing interventions, in selecting participants, in validating measurements, or in evaluating research activities (e.g., Astleitner, 2011).

General Theoretical Shortcomings in Educational Intervention Research

When having a closer look at the actual given quality of educational intervention research, then three major issues arise in respect to theoretical foundations.

First, educational interventions need theory sophistication instead of early theory replacement: Of course, theories are used more or less significantly in intervention research to formulate hypotheses, to undertake construct validation, or to cope with alternative explanations or methodological problems what should lead to an enriched knowledge about a subject area (e.g., Nipedal, Nesdale, & Killen, 2010). However, this kind of progress from enrichment is endangered, because, especially, in educational research, researchers have complained repeatedly about theories that they "seem to replace one another, rather than subsume, extend, or complement other theories" (DiSessa & Cobb, 2004, p. 79). There is probably less progress and less confidence in nomological networks than it could be, because of a resistance in integrating different theories or background variables that were found to be effective in other studies (e.g., Lynch, 1983). This resistance might come from the fact that it is scientifically more rewarding and often easier to develop and test new theories than to expand old theories to more complex, more dynamic, or more systemic ones.

Second, educational interventions need to apply methods for theoretical problem solving: Undoubtedly, the design of interventions should be well balanced in respect to the state of the art in methodological and theoretical developments. However, there is much more literature on applying scientific methods for gathering, processing, and validating data than on constructing theories and on integrating them in designing and evaluating intervention research (e.g., Reynolds, 2007). One might conclude that, within the field educational intervention research, methodological problem solving is much stronger implemented that theoretical problem solving. For example, although many have shown the functions of theories in the scientific process, they have not elaborated on how to use them systematically for the optimization of intervention research (e.g., Westermann, 2000). Others have focused in mixed approaches, both on (more quantitative) testing and on (more qualitative) developing intervention research, however, without establishing theoretically relevant criteria or procedures that can help when experiencing complex

research problems like, for example, when identifying moderating or irrelevant factors, or explaining unexpected data, or handling theory or method revision (e.g., Johnson & Christensen, 2004; Patry, 2011).

Third, not only measurements or assessments but also educational interventions need construct validation: There is a whole scientific "industry" dealing with the construction of tests for measuring personality characteristics and related variables (e.g., Embretson, 2007). However, within educational intervention research, theories have played a major role in construct validation of dependent variables, but not of independent ones. Bredenkamp (1979) has pointed out that within a construct validation of independent variables, it has to be asked whether an intervention is a representative for all possible interventions, for related internal processes, or for all modalities of a variable. In addition, Patry (1990) has shown that interventions can have both convergent and discriminant validity as an indicator of construct validity. Probably, this lack in construction validation comes from priorities given in educational intervention research. These priorities might range effectiveness or efficiency prior to theoretical foundation and development neglecting that theories are the key factor in generating, selecting, or validating intervention alternatives.

Implementing Theories and the Quality of Educational Interventions

Educational science and related disciplines of social research have not done anything to handle theses shortcomings. They have suggested meta-theoretical or general frameworks and processes of theory use in order to improve the theoretical quality of interventions.

Meta-theoretical or general frameworks. Evans, Meyer, and Buckley (2008) have proposed a multi-dimensional model for the design of any clinical treatment. According to this model, all interventions should a) manipulate immediate consequences of a behavior, b) re-arrange environments, c) facilitate alternative skills, and d) design long-term prevention through new behavior patterns. Also, Finney and Moos (1989) have presented a conceptual model of intervention design and evaluation. The model concerns three processes: intervention selection, intervention process, and the onset and course of the intervention problem. All processes are related to "theories of treatment" which specify "the techniques or procedures to be applied to bring about desired outcomes" (p. 310). Theories of treatment are assumed to improve interventions by getting more information about the internal mechanism and external factors that contribute to intervention success or failure. Both models were generated for presenting a general framework for increasing intervention effectiveness, however, they do not address in detail the role of theories for quality assurance in interventions. Lynch (1999) has pointed out that background factors related to external validity should be integrated as moderators into theories. Such suggestions represent meta-theoretical assumptions that can guide theory development in respect to environmental aspects of interventions. Some kind of meta-theoretical stimulation has also come from Hettinger Steiner and Carr (2003). They have tried to merge more or less isolated results from traditional research with new theoretical models (in the field of research on giftedness). This intense theory-research-interaction has resulted in exemplary ideas about the establishment of complex and dynamic system models that can deliver a more complete picture of human development and also related intervention research. They have not answered the question how theories can lead to better interventions, but what kind of theories are needed to increase the probability of more effective interventions.

Processes and areas of theory use. DiSessa and Cobb (2004) have described a process illustrating how theories can improve innovation in interventions (i.e., "design experiments"). Theories are shown in a process of generating, selecting, and validating intervention alternatives. Their presented case studies contain some examples of questions and criteria, which could be helpful for intervention design. Whyte (2006) has shown that treatment theories can help in selecting "active ingredients" of interventions, in the choice of participants, in finding outcome measures, or in study design and establishing comparison conditions. They have discussed the role of inclusion and exclusion criteria, limits of treatment utility, prognostic variables, treatment response modifiers, proximal and distal targets, and others. Siemonsma, Schröder, Roorda, and Lettinga (2010) have illustrated the use of treatment theories in rehabilitation medicine, also a field of intervention research. They have shown how the use of treatment theories changes the design of clinical trials. Treatment theories have allowed to a) describe the assumed process of intervention change and effects in more detail, b) state characteristics of participants (e.g., language skills) that could affect significantly intervention effects, c) design control groups with participants who could benefit most from interventions, and d) select outcome measures that allow to cover as many as possible effects of an intervention.

All these activities that show how to use theories to improve intervention research have identified some important goals, processes, and areas. However, they have not delivered some kind of tools that could assist designers and evaluators of educational interventions in a comprehensive and systematic way. Such a tool could consist of areas and related questions that can guide the use of theories in a step-by-step intervention design and evaluation. Valentine and Cooper (2008) have presented such a tool, i.e., a quality scale, but with a strong focus on the role of methods, but not on theories.

It is the goal of this paper, to develop and present such a tool that could assist in designing and evaluating educational interventions from a theoretical perspective. The tool, called ITQ (Intervention Theory Questions) should

be based on recent intervention research, distinguish different areas, and consist of easily applicable questions. Based on these questions, it will be illustrated - by an example in reducing aggressive behavior - how the design of an educational intervention changes in comparison to a situation in which these questions were not in use. Here, a bully prevention intervention from Ross and Horner (2009) was selected to illustrate the effects of using ITQ. This study was considered because it was well designed from a methodological and theoretical perspective and because it showed high effectiveness. It is not intended to criticize this study, but to illustrate how the goals, designs, and evaluations of intervention studies can change when using ITQ.

Design and Evaluation of an Educational Intervention without and with Using ITQ

Tables 1a to 1d contain questions that can stimulate decisions within intervention studies based on theoretical aspects. All questions of the ITQ should be used when designing and evaluating educational interventions.

Include Tables 1a to 1d about here!

These questions were found by scanning, paraphrasing, integrating, calibrating, and organizing many contributions of related methodological and theoretical literature dealing with research design (see Alisch, 1995; Barnett, Daly, Jones, & Lentz, 2004; Borsboom, Mellenbergh, & van Heerden, 2004; Bracht & Glass, 1968; Brewer & Hunter, 2006; Campbell, 1957; DiSessa & Cobb, 2004; Hagermoser Sanetti & Kratochwill, 2009; Highhouse, 2009; Johnson & Christensen, 2004; Kane, 1992, 2001; Kelly, Baek, Lesh, & Bannan-Ritland, 2008; Lipsey, 1990; Lissitz & Samuelsen, 2007; Lynch, 1983; McCaul & Glasgow, 1985; Meier, 2004; Mitchell & Jolley, 2010; Patry & Perrez, 2000; Reigeluth, 1999; Schurz, 2008; Shadish, Cook, & Campbell, 2002; Shadish, Cook, & Leviton, 1991; or Snow, 1987).

In order to identify and organize these questions, strategies for "drawing and verifying conclusions" like, for example, noting patterns, clustering, making contrasts/comparisons, or subsuming particulars into the general from qualitative data analysis methods were used (Miles & Huberman, 1994). First, contributions from literature were searched for assumptions about the integration of theories in intervention research. Second, these assumptions were organized in respect to their focus on situational parameters, subject characteristics, treatment design, mediating variables, and/or dependent variables. Third, assumptions were transformed into questions and re-organized into global areas (problem solution, theory development, intervention design and implementation contexts, and discussions; see Tables 1a to 1d). Finally, the categorization and formulation of questions were calibrated in order to avoid definitional inaccuracies and overlappings of meanings.

Having ITQ in mind, the planning, design, and evaluation of educational interventions can change considerably. There are two main ways how to use the ITQ. One way can be described as being related to a more exploratory use with less specific focuses. That means that a researcher uses the ITQ to find some stimulating ideas for decisions about different aspects of educational interventions. An unstructured scanning of the ITQ without specific goals drives this process. Another way is a more systematic use of ITQ with a clear focus. The researcher has a certain goal (e.g., improving the effectiveness of an intervention) and this goal should be achieved by selecting many or all questions of a specific ITQ area (e.g., effectiveness). Of course, these two ways can be mixed resulting in a scanning-focusing procedure.

The following section should exemplarily show how the design and evaluation of the Ross and Horner (2009) intervention changes when using the ITQ in different ways. For each global area of the ITQ, questions were randomly selected out of the ITQ and answered by considering additional literature.

Problem. Ross and Horner (2009) concluded from their review of existing bully prevention interventions that results have been weak and mixed with high variations, even with negative effects (for example, when students are teaching each other anti-bullying). Another problem has come from the fact that many bully prevention studies have been based on self-reports without considering systematic observations.

Ross and Horner (2009) have identified more or less methodological reasons (e.g., inconsistencies in evidence) related to their major research problem. However, when using ITQ, the questions out of the global area of problem solution are relevant (see Table 1a). These questions should lead to a complex problem view by searching, comparing, evaluating, and/or integrating different theories. For example, Lindsay and Anderson (2000) have developed a "general affective aggression model" that could lead together with the "social-cognitive information-processing model" from Boxer and Dubow (2002) to a set of sophisticated research problems for educational interventions. Ross and Horner (2009) have theoretically clearly focused on behavioral aspects of the aggression problem (e.g., the role of social rewards). They have not focused in detail on "social-cognitive information-processing" in which, for example, the attention and interpretation of external cues or beliefs that support the use of aggression are important. However, when using the ITQ questions and when having an expanded social-cognitive theoretical perspective, than, for example, the following research questions might arise: (1) Should anti-aggression trainings not only change behavior problems, but also problem-related social-cognitive information processing (see

findings problems in the exploration area in Table 1a)? (2) Can social-cognitive information processing be measured or tested successfully in certain, for example, stressful aggression settings (see the criteria of empirical testability as part of the theory evaluation area)? (3) Is social-cognitive information processing part of a mediating mechanism that could help to reduce aggressive behavior (see mediating processes within the goal building area)? (4) Are there consistent (e.g., only positive) or inconsistent (e.g., positive and negative) theoretically assumed effects, when a behavioral and a social-cognitive perspective are integrated (see integration within triangulation area)? (5) Could a social-cognitive perspective realize a programmatic intervention progress in respect to behavioral educational interventions on aggression reduction (see treatment development within progress area)?

Theoretical background. The theoretical background of the study from Ross and Horner (2009) was based on definitions of bullying and on causal variables that are assumed to influence bullying. Especially, variables in a behavioral approach were focused over which parents and others had control and which included events that precede and follow bullying behavior (like peer attention). It was hypothesized that there is a relationship between bully prevention and physical and verbal aggression on playground activities. Prevention should increase the probability that victims will walk out of a problem situation and decrease social rewarding of problem behavior. In addition, prevention should increase stopping and helping victim behaviors of bystanders.

Of course, theories can be used to find important research problems, but their main task is to deliver the theoretical background for educational interventions. This background leads to the formulation of hypotheses that are related to a more or less causal relationship between independent and dependent variables. Within ITQ and its global area of theory development, there are such questions (see Table 1b). One such question is, for example, related to the linear or non-linear type of relationship between variables (see covariation area). Within literature, meta-analyses have shown that similar interventions for reducing aggression can produce positive but also negative effects that are probably indicating an up-and-down-process (e.g., Merrell, Gueldner, Ross, & Isava, 2008). For example, entering into a dialog about aggressive behavior can produce evaluations of persons resulting in negative emotions or reactance. Such negative evaluations can increase aggressive behavior instead of decreasing it. So, there might be an increase in the short term and probably a decrease in the long run when, for example, higher order cognitions (e.g., assessment of outcomes) arise, resulting in an overall non-linear inverse U-term-relationship.

When such a theoretical analysis reveals a non-linear relationship between independent and dependent variables, then the designs of the intervention and of measurements are different in comparison to a linear setting. Ross and Horner (2009) have used a three-tiered prevention model in their intervention that can handle no or negative intervention effects step-by-step and that allow to measure no or negative results for a significant duration of time. So, they have important ingredients to handle non-linear relationships between their interventions and dependent variables, although no significant theoretical elaborations can be found why students do not respond positively to their interventions. Such elaborations could, for example, be based on theoretical approaches dealing with the development and experience of aggression and using concepts like the stability of causes for aggression (e.g., Canary, Spitzberg, & Semic, 1998): During an intervention, people´s attributions or explanations why an aggressive act occurs could change over time and such a change should be reflected theoretically.

Another theoretical aspect could, for example, concern hierarchical classes, i.e., the question whether a stepwise change in the dependent variable can be related to a stepwise change in independent variables (see hierarchical [latent] classes within the specific area of process explanation in Table 1b). Such a theory could increase validity because it describes developments as transitions between latent classes (see Borsboom, Mellenbergh, & van Herden, 2004). In respect to aggression, different hierarchically organized development models could be found: So, for example, the affective goal taxonomy by Krathwohl, Bloom and Masia (1964) using concepts like attending (awareness, willingness to receive, controlled or selected attention), responding (acquiescence in responding, willingness to respond, satisfaction in response), valuing (acceptance of a value, preference of a value, commitment), organization (conceptualization of a value, organization of a value system), or characterization by a value or value complex (generalized set, characterization). Other relevant theoretical approaches could come, for example, from hierarchically organized stages of moral development by Kohlberg (2001). Within the intervention study from Ross and Horner (2009), such a hierarchical model could, for example, be linked with a stepwise intervention procedure that is based on a "changing criterion design": In a first step, attending to aggressive behavior and related criteria could be achieved by the intervention; when the first criteria is meet then responding to aggressive behavior and related criteria should be in the focus of intervention and so on.

Intervention design and implementation contexts. The Ross and Horner (2009) intervention "BP-PBS (Bully Prevention in Positive Behavior Support)" was designed to fit within a system of schoolwide behavior support for elementary-grade students, to implement respectfulness in a school, to minimize social reinforcement of problem behavior by a three-step response (stop-walk-talk) pattern, to pre-correct three-step response before acting, to teach good replying when the three-step response is applied, and to train staff in handling student reports of problem behavior. The BP-PBS-intervention was implemented in three elementary schools in an overall period of 60 school

days together with staff training and assessing interobserver agreement and social validity. In each school, two students with high levels of aggression participated in a multiple baseline design study. Measuring student knowledge and staff adherence to intervention components assessed the fidelity of implementation. Frequency of physical and verbal aggression during lunch recess, victim responses to problem behavior, and social responses from bystanders were recorded as dependent variables.

When designing and evaluating an intervention, an important question is, for example, the one of the dosage in respect to invested effort in time, personal and so on (see specific area of active ingredients in Table 1c). A too high dosage decreases efficiency, a too low dosage increases the probability of ineffectiveness. Like, within the study of Ross and Horner (2009), many educational interventions range from a few hours to few days duration of time. Such durations as a main indicator of the dosage of an intervention are often based on former experiences without using significant theoretical foundations or discussions. Such theoretical foundations should optimize intervention effects and assist in answering questions about when an intervention effect starts, how long it lasts, or whether there is a "treatment-effect decay" (for example, Lipsey, 1990, p. 156 distinguishes four forms of such decays: immediate effect, delayed effect, immediate effect and rapid decay, or early effect and slow decay). Such questions can easily be answered after an intervention when having descriptive process data about intervention effects. However, in order to explain, predict, and optimize the intervention, theories have to be used with a developmental perspective (e.g., Frick & White, 2008). A developmental perspective on a short-term micro- and a long-time macro-focus can help in getting knowledge for intervention design in respect to, for example, the stability of effects (based on situational influences and personality traits), to stages of different levels of special reactivity or sensitivity to interventions, to factors that protect or disturb developments and so on. Handling these questions might have also improved the effects from the Ross and Horner (2009) study, especially in respect to the duration and the supporting settings of interventions. For example, when there is a theory indicating that intervention effects are to some degree based on novelty effects and that novelty effects disappear within a certain period of time, then additional intervention efforts (e.g., motivational design applications like implementing challenging performance objectives; see van Merrienboer & Kirschner, 2007) should be made before the novelty effect decreases significantly.

In Table 1c, within the specific area of passive ingredients, there is a focus on acceptance design dealing with the question whether the intervention could produce reactance in participants. The intervention from Ross and Horner (2009) utilizes a comprehensive schoolwide program to prevent aggressive behavior that reduces freedom of acting for students significantly what could result in reactance. Reactance occurs when the possibility to choose behavior is restricted. It corresponds with a motivational and also aggressive arousal together with the goal of reinstalling the freedom to choose. From a theoretical point of view, reactance can be reduced when, for example, freedom is reinstalled directly or indirectly (e.g., by allowing some kind of not goal relevant behavior in other settings under certain conditions) (Dickenberger, 2006). In the study from Ross and Horner (2009), reactance of participants could be one explanation why problem behavior was not eliminated completely. Therefore, handling reactance to the intervention could be part of supplemental intervention through individual support plans in order to achieve additional reductions in aggressive behavior.

Discussions. Reductions of students´ problem behaviors ranged from 53 to 86 % from baseline means with steadily increasing trends after fully implementing BP-PBS in the Ross and Horner (2009) study. The intervention was also associated with increases in victim and bystander responses to problem behavior. Assessments of intervention fidelity showed that students were able to learn and that staff was able to implement the BP-PBS curriculum with high satisfaction. Within discussions, the none-use of bullying language, problems in implementation efficiency and long-term intervention perspectives, and the establishment of positive behavior support at schools as a pre-condition of BP-PBS-interventions were addressed. As limitations of the study, Ross and Horner (2009) have mentioned, that some type of students might not respond to BP-PBS-interventions, that problem behavior was not eliminated completely, and that the collected data might not be representative for student behavior in other school settings.

Within discussions, usually, results of educational interventions are evaluated based on theoretical, methodological, and practical aspects. Such evaluations result in suggestions for future educational intervention activities. ITQ offers in its discussions section different specific areas that could stimulate such evaluations (see Table 1d).

For example, one important question could be whether educational interventions have different effects for different people (based on different sex, age, personality characteristics etc.) and why these different effects could occur (see trait-intervention effect in the specific area of effectiveness, Table 1d). There is, for example, a considerable inconsistency in findings testing aggression x gender effects: From some studies, it could be concluded that girls experience aggression and act different from boys, or that girls have different aggression effecting contexts in comparison to boys; other studies did not find such differences (see, for example, Martino, Ellickson, Klein, McCaffrey, & Orlando Edelen, 2008). Boys might react different to group norms or institutional norms in

comparison to girls; they might have different aggressive intentions or self-regulation skills and so on (see, for example, Nipedal, Nesdale, & Killen, 2010). Despite the inconsistencies in research findings, in respect to educational intervention focusing on aggression, it has to be explained whether interventions should be different for boys and girls (and other personality characteristics) or not. In the study from Ross and Horner (2009), there are no explicit hints that interventions were designed differently for girls and boys or for other personality characteristics. However, such a different design could be helpful, especially for participants that did not respond to their interventions.

Another question from ITQ can come from, for example, focusing on side effects (see specific areas in Table 1d). Side effects are unintended positive or negative outcomes of an educational intervention. Coping with side effects is a central issue in, for example, medical or psychological therapy. One way of handling side effects is to measure them during and after interventions, another way is to explore them by using theories. After having some theoretical and empirical evidence of side effects, they can be handled in future educational interventions. In order to answer the questions of possible positive or negative side effects of an intervention, theories that cover peripheral phenomenon can be used. In respect to the study from Ross and Horner (2009), it has, for example, to be asked whether an intervention for reducing aggressive behavior has covert, indirect, or improbable effects or whether effects produce insidious long-term developments or cascades resp. combinations of unintended effects. For example, Nipedal, Nesdale, and Killen (2010) distinguished between overt aggression (like hitting or pushing), but also subtle ones (like deceptions or manipulation). It might be possible that controlling or reducing overt aggression could produce - as substitution - forms of aggression that are not easily to observe or measure. Or, according to Caprara, Barbaranelli, Pastorelli, Bandura, and Zimbardo (2000), reducing aggression and increasing prosocial behavior correspond with changes in academic achievements. Such improbable effects at first sight could be used, for example, in designing combined aggression reduction and school improvement programs. Such programs could deal with both social- and achievement-related developments at the same time. Focusing on side effects could help to find such innovations.

Discussion

The purpose of this paper was to present a tool, the ITQ, that should assist researchers and practitioners when planning, designing, and evaluating educational interventions in respect to theoretical foundations.

Of course, this tool is not perfect. The resulting questions represent a collection of questions that are neither fully exclusive nor saturated. On the one hand, some questions could be classified in other areas or are strongly related to other questions. On the other hand, questions do not cover all possible theoretical issues related to intervention research.

The questions do also not have a common and consistent theoretical, meta-theoretical, or methodological framework. However, they deliver a research tool that should assist intervention designers and evaluators in their work. Questions should lead to invention research activities with expanded theoretical foundation, increased reliability and validity, and improved effectiveness. Especially, they also should lead to come to a multi-faceted view of problem behaviors, stimulate innovative intervention solutions, assist in reflecting genuine methodological questions from a widened theoretical perspective, or help in finding explanations for unexpected effects of interventions.

There is also no clear plan or program, when to use these questions. Questions of the ITQ should be asked at all stages of educational interventions. For example, questions concerning discussions should not be asked after the intervention when conceptualizing the discussion part of a report, but also especially during planning an intervention.

In addition, empirical evidence or further experiences concerning the ITQ are missing that could help in evaluating how ITQ can change educational interventions.

ITQ should be a starting point for establishing a more sophisticated focus on theoretical backgrounds for educational interventions.

References

Alisch, L.-M. (1995). Grundlagenanalyse der Pädagogik als strenge praktische Wissenschaft [Analysing the basics of education as a strict practical science]. Berlin: Dunker & Humblot.

Astleitner, H. (2011). Theorieentwicklung für SozialwissenschaftlerInnen [Theory building methods for social scientists]. Wien: Böhlau, UTB.

Barnett, D. W., Daly III, E. J., Jones, K. M., & Lentz, Jr., F. E. (2004). Response to intervention: Empirically based special service decisions from single-case designs of increasing and decreasing intensity. The Journal of Special Education, 38, 66-79.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. Psychological Review, 111, 1061-1071.

Boxer, P. & Dubow, E. F. (2002). A social-cognitive information-processing model for school-based aggression reduction and prevention programs: Issues for research and practice. Applied & Preventive Psychology, 10, 177-192.

Bracht, G. H. & Glass, G. V. (1968). The external validity of experiments. American Educational Research Journal, 5, 437-474.

Bredenkamp, J. (1979). Das Problem der externen Validität pädagogisch-psychologischer Untersuchungen [The problem of external validity in studies in the field of Educational Psychology]. In J. Brandstätter, G. Reinert & K. A. Schneewind (Hrsg.), Pädagogische Psychologie. Probleme und Perspektiven (S. 267-289). Stuttgart: Klett-Cotta.

Brewer, J. & Hunter, A. (2006). Foundations of multimethod research. Synthesizing styles. Thousand Oaks, CA: Sage.

Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. Psychological Bulletin, 54, 297-312.

Canary, D. J., Spitzberg, B. H. & Semic, B. A. (1998). The experience and expression of anger in interpersonal settings. In P. A. Andersen & L. K. Guerrero (Eds.), Handbook of communication and emotion (pp. 189-213). San Diego, CA: Academic Press.

Caprara, G. V., Barbaranelli, C., Pastorelli, C., Bandura, A., & Zimbardo, P. G. (2000). Prosocial foundations of children´s academic achievement. Psychological Science, 11, 302-306.

Dickenberger, D. (2006). Reaktanz [Reactance]. In H.-W. Bierhoff & D. Frey (Hrsg.), Handbuch der Sozialpsychologie und Kommunikationspsychologie (S. 96-102). Göttingen: Hogrefe.

DiSessa, A. A. & Cobb, P. (2004). Ontological innovation and the role of theory in design experiments. The Journal of the Learning Sciences, 13, 77-103.

Embretson, S. E. (2007). Construct validity: A universal validity system or just another test evaluation procedure? Educational Researcher, 36, 449-455.

Evans, I. M., Meyer, L. H. & Buckley, L. (2008). Interpreting the evidence on interventions with challenging behaviour: Theoretical, contextual, and cultural perspectives from New Zealand. Australian Journal of Rehabilitation Counselling, 14, 79-93.

Finney, J. W. & Moos, R. H. (1989). Theory and method in treatment evaluation. Evaluation and Program Planning, 12, 307-316.

Frick, P. J. & White, S. F. (2008). The importance of callous-unemotional traits for developmental models of aggressive and antisocial behavior. Journal of Child Psychology and Psychiatry, 49, 359-375.

Fuchs, L. S., Fuchs, D., & Speece, D. L. (2002). Treatment validity as a unifying construct for identifying learning disabilities. Learning Disability Quarterly, 25, 33-45.

Hagermoser Sanetti, L. M. & Kratochwill, T. R. (2009). Toward developing a science of treatment integrity: Introduction to the special series. School Psychology Review, 38, 445-459.

Hettinger Steiner, H. & Carr, M. (2003). Cognitive development in gifted children: Toward a more precise understanding of emerging differences in intelligence. Educational Psychology Review, 15, 215-246.

Highhouse, S. (2009). Designing experiments that generalize. Organizational Research Methods, 12, 554-566.

Johnson, B. & Christensen, L. (2004). Educational research. Quantitative, qualitative, and mixed approaches (2nd ed.). Boston: Pearson.

Kane, M. T. (1992). An argument-based approach on validity. Psychological Bulletin, 112, 527-535.

Kane, M. T. (2001). Current concerns in validity theory. Journal of Educational Measurement, 38, 319-342.

Kelly, A. E., Baek, J. Y., Lesh, R. A., & Bannan-Ritland, B. (2008). Enabling innovations in education and systematizing their impact. In A. E. Kelly, R. A. Lesh, & J. Y. Baek (Eds.), Handbook of design research methods in education. Innovations in science, technology, engineering, and mathematics learning and teaching (pp. 3-18). New York, London: Routledge.

Kohlberg, L. (2001). Moralstufen und Moralerwerb. Der kognitiv-entwicklungstheoretische Ansatz (1976) [Moral steps and acquisition. The cognitive-developmental approach (1976)]. In W.

Edelstein, F. Oser & P. Schuster (Hrsg.), Moralische Erziehung in der Schule. Entwicklungspsychologie und pädagogische Praxis (S. 35-61). Weinheim: Beltz.

Krathwohl, D. R., Bloom, B. S. & Masia, B. B. (1964). Taxonomy of educational objectives: The classification of educational goals. Handbook II: Affective domain. New York: Longman.

Lindsay, J. J. & Anderson, C. A. (2000). From antecedent conditions to violent actions: A general affective aggression model. Personality and Social Psychology Bulletin, 26, 533-547.

Lipsey, M. W. (1990). Design sensitivity. Statistical power for experimental research. Newbury Park, CA: Sage.

Lissitz, R. W. & Samuelsen, K. (2007). A suggested change in terminology and emphasis regarding validity and education. Educational Researcher, 36, 437-448.

Lynch, J. G. Jr. (1983). The role of external validity in theoretical research. Journal of Consumer Research, 10, 109-111.

Lynch, J. G. Jr. (1999). Theory and external validity. Journal of the Academy of Marketing Science, 27, 367-376.

Martino, S. C., Ellickson, P. L., Klein, D. J., McCaffrey, D., & Orlando Edelen, M. (2008). Multiple trajectories of physical aggression among adolescent boys and girls. Aggressive Behavior, 34, 61-75.

McCaul, K. D. & Glasgow, R. E. (1985). Preventing adolescent smoking: What have we learned about treatment construct validity? Health Psychology, 4, 361-387.

Meier, S. T. (2004). Improving design sensitivity through intervention-sensitive measures. American Journal of Evaluation, 25, 321-334.

Merrell, K. W., Gueldner, B. A., Ross, S. W., & Isava, D. M. (2008). How effective are school bullying intervention programs? A meta-analysis of intervention research. School Psychology Quarterly, 23, 26-42.

Miles, M. B. & Huberman, A. M. (1994). An expanded sourcebook qualitative data analysis (2nd ed.). Thousand Oaks, CA: Sage.

Mitchell, M. L. & Jolley, J. M. (2010). Research design explained (7th ed.). Belmont, CA: Wadsworth.

Nipedal, C., Nesdale, D., & Killen, M. (2010). Social group norms, school norms, and children´s aggressive intentions. Aggressive Behavior, 36, 195-204.

Patry, J.-L. (1990). Feldforschung [Field research]. In L. Kruse, C.-F. Graumann, & E.-D. Lantermann (Hrsg.), Ökologische Psychologie (S. 183-195). München: Psychologie Verlags Union.

Patry, J.-L. (2011). Methodological consequences of situation specificity: Biases in assessments. Frontiers in Psychology. Quantitative Psychology and Measurement, 2, Article 18, 1-17.

Patry, J.-L. & Perrez, M. (2000). Theorie-Praxis-Probleme und die Evaluation von Interventionsprogrammen [Theory-practice problems and the evaluation of intervention programs]. In W. Hager, J.-L. Patry, & H. Brezing (Hrsg.), Evaluation psychologischer Interventionsmaßnahmen (S. 19-40). Bern: Huber.

Pourret, O., Naim, P., & Marcot, B. (2008). Bayesian networks. A practical guide to applications. Chichester: Wiley.

Reigeluth, C. M. (1999). What is instructional-design theory and how is it changing? In C. M. Reigeluth (Ed.), Instructional-design theories and models. A new paradigm of instructional theory. Vol. II (pp. 5-29). Mahwah, NJ: Erlbaum.

Reynolds, P. D. (2007). A primer in theory construction. Boston: Pearson.

Ross, S. W. & Horner, R. H. (2009). Bully prevention in positive behavior support. Journal of Applied Behavior Analysis, 42, 747-759.

Schurz, G. (2008). Einführung in the Wissenschaftstheorie [Introduction to the theory of science] (2. durchges. Aufl.). Darmstadt: Wissenschaftliche Buchgesellschaft.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin Comp.

Shadish, W. R. Jr., Cook, T. D., & Leviton, L. C. (1991). Foundations of program evaluation. Theories of practice. Newbury Park, CA: Sage.

Siemonsma, P. C., Schröder, C. D., Roorda, L. D., & Lettinga, A. T. (2010). Benefits of treatment theory in the design of explanatory trials: Cognitive treatment of illness perceptions in chronic low back pain rehabilitation as an illustrative example. Journal of Rehabilitation Medicine, 42, 111-116.

Snow, R. E. (1987). Aptitude complexes. In R. E. Snow & M. J. Farr (Eds.), Aptitude, learning, and instruction. Vol. 3: Conative and affective process analyses (pp. 11-34). Hillsdale, NJ: Erlbaum.

Valentine, J. C. & Cooper, H. (2008). A systematic and transparent approach for assessing the methodological quality of intervention effectiveness research: The study design and implementation assessment device (Study DIAD). Psychological Methods, 13, 130-149.

van Merrienboer, J. J. G. & Kirschner, P. A. (2007). Ten steps to complex learning. Mahwah, NJ: Erlbaum.

Weber, S. J. & Cook, T. D. (1972). Subject effects in laboratory research: An examination of subject roles, demand characteristics, and valid inference. Psychological Bulletin, 77, 273-295.

Westermann, R. (2000). Wissenschaftstheorie and Experimentalmethodik [Theory of science and methods of experimental research]. Göttingen: Hogrefe.

Whyte, J. (2006). Using treatment theories to refine the designs of brain injury rehabilitation treatment effectiveness studies. Journal of Head Trauma Rehabilitation, 21, 99-106.

Yeaton, W. H. & Sechrest, L. (1981). Critical dimensions in the choice and maintenance of successful treatments: Strength, integrity, and effectiveness. Journal of Consulting and Clinical Psychology, 49, 156-167.

Table 1a
Questions Related to Treatment Theories in Intervention Research

Global Areas	Specific Areas	Questions
Problem solution	Exploration	Finding problems: Which problems should be in focus of intervention research activities from a theoretical point of view? Finding solutions: Which intervention solutions to a problem can be suggested based on theoretical assumptions?
	Goal building	Causal relationships: Why should causal relationships between dependent and independent variables be in the focus of intervention research for theoretical reasons? Mediating mechanism: Why should mediating processes between dependent and independent variables be in the focus of research due to theoretical arguments? Optimization of interventions: Why should there be a focus on the goal of optimization of interventions due to theoretical reasons? Standard setting: What low and high levels of intervention goals can be fixed in relation to theoretical assumptions or evaluations? Impact: Why should there be an increased focus on the impact of interventions from a theoretical perspective?
	Triangulation	Multiple perspectives: What other theories confirm or disconfirm the theoretical background of an intervention? Integration: Can different theories be combined in order to build a consistent basis for intervention activities?
	Progress	Treatment development: How can a long-term improvement of interventions be achieved? What represents programmatic intervention progress from a theoretical perspective? Theory development: What represents long-term progress in theories for interventions from a (meta-)theoretical point of view?
	Theory evaluation	Primary quality standards: Does a given theory for interventions deliver clear concepts, logical coherence, empirical testability, and probable effectiveness, especially also when confronted with other theories? Secondary quality standards: Does a given theory for interventions realize parsimony, high integrative power, or stimulation of other research activities, especially when confronted with other theories?

Table 1b
Questions Related to Treatment Theories in Intervention Research

Theory development	Covariation	Correlation: Do, from a theoretical point of view, independent and dependent variables occur together? Does the independent variable occur before a change in the dependent variables? Type of relationship: Is there a linear or a non-linear (theoretical) relationship between independent and dependent variables?
	Other causes eliminated	Control variables: Do other variables affect the relationship between independent and dependent variables? How can these other variables be handled due to theoretical assumptions? Missing mediators: Which potential variables have not been considered as mediators between dependent and independent variables? What theoretical reasons are given for their exclusion? Irrelevant factors: Which variables are theoretically assumed as not being effective in given effect patterns of independent and dependent variables?
	Process explanation	Hierarchical (latent) classes: Is it possible to describe theoretically a stepwise change in dependent variables based on a stepwise change in independent and mediating variables? Can these steps hierarchically be organized out of theoretical assumptions? Sequence of effects: Allow theoretical assumptions to state the dynamics of mediating processes? Do some mediating processes occur before, simultaneously, or after others? Manipulation checks: How and when should tests be used for measuring the effects of independent variables on mediating processes according to theoretical assumptions?
	Contrasting conditions	Non-effective or reverse conditions: Under which (theoretically assumed) circumstances, independent variables and/or mediating processes could have no or even negative effects on dependent variables?

Table 1c
Questions Related to Treatment Theories in Intervention Research

Intervention design and implementation contexts	Active ingredients	Activities, episodes, materials: Are there theories that could help for selecting and combining the core elements of an intervention? Manipulate consequences, arrange environments, facilitate alternative skills, and make preventive actions: Allow theoretical assumptions to define and design the embedding of the core elements of the intervention? Dosage: Are there theoretical assumptions that can help in deciding how strong an intervention should be implemented? Are there theoretical suggestions for the duration, repetition, etc. of an intervention?
	Passive ingredients	Perceptual design: Do theories allow to evaluate whether the intervention can reach the participants from a perceptual perspective? For example, can participants hear and/or see all important elements in the whole process of the intervention? Communication design: From a theoretical point of view: Is there a use of language within the intervention that can be understood by all participants? Acceptance design: Is the intervention designed in a way that can be accepted by the participants from an ethical, attitudinal, or emotional point of view? Is there some element in the intervention that could produce resistance or reactivity?
	Selection of participants	Are there theoretical reasons why some participants should take or should not take part in the intervention?
	Possible modalities	Do the design of the intervention cover all possible ways for implementing the independent variable? Are there other more or less effective ways from a theoretical perspective?
	Bias handling	Integrity: Do theories help to evaluate whether the intervention can be delivered as intended or not? Hawthorne effect: Are there theoretical reasons why knowledge of being part of an intervention can change behavior? Novelty effect: Do, from a theoretical point of view, conditions in the intervention be effective or non-effective because they are new to the participants? Disruption effect: Do, from a theoretical point of view, conditions in the intervention be effective or non-effective because they are unfamiliar to the participants?

Table 1d
Questions Related to Treatment Theories in Intervention Research

Discussions	Effectiveness	Effect size: How strong can and should be the effect of an intervention due to theoretical reasons? Efficiency: Is there a theoretical chance to increase the efficiency (input-output-relationship) of an intervention? Robustness: How likely is it from a theoretical perspective that an intervention can reproduce its effects in future attempts? Selection problem: What does it mean for intervention and theory effectiveness when only parts of a theory are used in intervention research? Package: When multiple independent variables are combined to intervention packages, how do these variables contribute to effectiveness from a theoretical perspective? What independent variables will produce ultimate, instrumental, or intermediate effects during intervention? Complementary effects: Do different independent variables in intervention packages assist each other in being effective based on theoretical assumptions? Additive/mixed effect: Do different independent variables in intervention packages depend (theoretically) in their effectiveness on other independent variables or not? Do independent variables have an additional or mixed effect in respect to other independent variables? Trait-intervention effect: Do interventions have different effects for different people due to theoretical assumptions? According to theory, are mediating processes different for different people? Are there theoretical assumptions that show how variables work differently in different people?
	Side effects	How likely is it that an intervention has unintended positive or negative effects? How can such effects be explained by theories?
	Follow ups	What theoretical assumptions explain why interventions are effective or not effective after different (short or long) periods of time? Which mediating processes produce or reduce effects after the intervention?
	Other contexts	Other situations: Are there theoretical reasons that allow to evaluate whether an intervention is effective in different situations? Could an intervention effectively be used in daily life situations from a theoretical point of view? Other people: Could an intervention (theoretically) be effective for people that are different from the participants in the intervention?