Fostering Evidence-based Practice and Argument Evaluation in Social Work

---

Dissertation zum Erwerb des Doctor of Philosophy (Ph.D.) am Munich Center of the Learning Sciences der Ludwig-Maximilians-Universität München

Vorgelegt von

Florian Spensberger

aus München

17.12.2019

1st Supervisor/ Erstgutachter:	Prof. Dr. Sabine Pankofer
2nd Supervisor/ Zweitgutachter:	Prof. Dr. Ingo Kollar
Tag der mündlichen Prüfung:	10.02.2020

Acknowledgements

The completion of this doctoral thesis would not have been possible without the support of various persons.

First, I would like to thank my grandmother who constantly emphasized the importance and the beauty of education. I know you would have loved to see this PhD becoming reality.

Second, I want to thank my dearly beloved wife Sandra, who put her goals on second place for years to make mine come true. Thank you for all your support and especially for all your love and efforts for our wonderful children Maja and Moritz. I also want to thank my father in law, Erich. You have been my role model for many years now. Without your silent guidance, I might not have developed the endurance and diligence that were so helpful throughout the last years.

Third, I want to thank my supervisors. Sabine, your constant support as well as your honest interest and passion for the research projects of this dissertation helped me to stay focused and motivated throughout the last three years. Furthermore, your overview and guidance were key for thought-out planning and respectively the completion of the dissertation within the expected timeframe. I also want to thank Ingo. For your openness regarding my different disciplinary background as well as your constant, patient, focused and detailed feedback. Both, the methodological quality of the studies of this dissertation as well as my personal learning progress would not have been the same without you. My deep gratitude goes also to Eileen. Thank you for your wonderful hospitality during my incubator stay. You offered me way more time than I could have asked for and thinking back to my time in California and to all our discussions

puts always a smile on my face. Thank you for sharing your wisdom with me. I am more than grateful to have you as a mentor and a friend. I also want to thank Frank Fischer. Achieving a PhD is never an easy task. Additionally, there are currently some structural barriers for social workers in Germany in that regard. Hence, without you and the opportunities you've provided me with REASON, achieving a PhD might very well have remained an unfulfilled dream for me.

Fourth, thank you Chris! For pointing my attention to the REASON program and for many inspiring talks. I have been enjoying our mutual work and I'm looking forward for more collaborations to come.

Fifth, I want to thank all the social work students who participated in the second empirical study. Thank you!

Finally, I want to thank all my fellow REASON PhD students for being such a great and supportive team. You guys rock! I am thankful for the friendships that have developed over the last three years, some of which might last for a long time. I also appreciated all the mutual support. Special thanks go out to Sarah, April and Arianne. Sarah, thank you for being so inclusive from day one. Coming from a quite different methodological background, your interest in other disciplines helped me to overcome my methodological insecurities and to align myself to my new surroundings. I also learned quite a bit during our endless hours of mutual coding. Thanks! April, thank you for editing and for making me film the video abstract! Arianne, thanks for showing me the beauty of statistics and open science as well as for editing and countless helpful and fun conversations. Cheers bro!

The research presented in this work was supported by the Elite Network of Bavaria [Project number: K-GS-2012-209]. I would like to extend my sincere gratitude to

the opportunities (conference attendance, incubator stay, international knowledge exchange) made possible by the ENB.

"Although I do not know for certain that less than optimal instruction is harmful to social work students, I think it is reasonable to assume that poorly trained social workers might have the potential to be harmful to the eventual recipients of their services.
Thus, I assert that social work education, like any other social work practice, has an ethical duty to work toward evidence-based social work education."

Joanne Yaffe (2013, p. 525)
Editor-in-chief, Journal of Social Work Education

Table of Contents

Part III: General Discussion, Implications and Conclusion

Preamble

Study 1 was submitted to the Journal 'Research on Social Work Practice' as an article entitled "How to teach evidence-based practice in social work: A systematic review". (https://journals.sagepub.com/doi/full/10.1177/1049731519852150).

Relevant documents regarding the studies' compliance with the General Data Protection Rules, the search history of the database searches, the data base of the retrieved articles, coding schemes as well as data of double coding and from the quality ratings of the articles were uploaded to the Open Science Framework. All this information is retrievable from *osf.io/dyubx*

Study 2 was submitted to the Journal of Social Work Education as an article entitled "Effects of worked examples and external scripts on fallacy recognition skills: A randomized controlled trial" and is currently under review.

Relevant documents regarding the studies' compliance with the General Data Protection Rules, pictures of the learning environment, the worked examples and the external script, the vignettes, the video on argumentation and criteria for good argumentation, the domain-specific prior knowledge test, the coding scheme, a log file of the coding process, the codebook and code for statistical analyses was uploaded to the Open Science Framework. Likewise, all collected data were anonymized and uploaded to the Open Science Framework. All this information is retrievable from *osf.io/5p86x*

The two articles are first-authored by Florian Spensberger. He wrote all drafts on which the co-authors gave feedback. Florian Spensberger used this feedback to change some arguments within the texts.

Part I

Introduction to Evidence-based Practice, Fallacies, Instructional Approaches and Scaffolds[1]

[1] According to the study regulations (§ 17), Part I can be considered as „ausführliche Zusammenfassung“ which is complemented by Part III.

Abstract

Bridging the research-practice gap is an endeavor that is shared by many helping professions such as medicine (Evidence-based Medicine Working Group, 1992) or psychology (Task Force, 1993). This is mainly due to the assumption that grounding practice on knowledge from empirical research may contribute to generating outcomes that are valued by clients (Gambrill, 2013; Task Force, 1993). Disciplinary social work is concerned with the question how to put research knowledge to practice at least since the 1970s (Jayaratne & Levy, 1979; Thyer & Myers, 2011). Okpych and Yu (2014) argue in their historical analysis that Evidence-based Practice is the most recent attempt to infuse social work practice with knowledge generated from empirical research. Two notions of Evidence-based Practice are prevalent in social work, one is called the process of Evidence-based Practice (EBP; Evidence-based Medicine Working Group, 1992; Gambrill, 1999; Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996) and another one Empirically Supported Treatments (ESTs; Task Force, 1993; Thyer & Myers, 2011). EBP refers to a decision-making process, ESTs to interventions that have been empirically evaluated. However, the implementation of Evidence-based Practice in social work practice still lags for various reasons. Some of these reasons such as lack of knowledge or diverse views of Evidence-based Practice are related to insufficient education of Evidence-based Practice in social work (e.g., Gray, Joy, Plath, & Webb, 2012; Scurlock-Evans & Upton, 2015). Furthermore, critical thinking skills such as the correct identification and explanation of fallacies (i.e., fallacy recognition skills) are considered a prerequisite for Evidence-based Practice (Aglen, 2016; Gambrill, 2019). Accordingly, numerous calls to facilitate social workers' fallacy recognition skills have been made (e.g.,

Gambrill & Gibbs, 2017). Yet, little is known on how to facilitate such skills within social work decision-making contexts.

The aims of this dissertation are to (1) provide an overview of the current body of research on how to facilitate EBP and the use of ESTs and to (2) assess this body of research with regard to the question how to best teach these approaches in social work. Another goal of the thesis is to (3) test the effectiveness of scaffolds, namely worked examples and an external script, within a case-based digital learning environment on social work students' fallacy recognition skills, which are an important prerequisite for EBP. Respectively, two empirical studies are conducted. Study one tackles aims number one and two, study two addresses the third goal.

The *first study* included in this doctoral thesis is a systematic review regarding the question how to teach Evidence-based Practice in social work. The narrative synthesis of the review reveals that the current body of research on education of Evidence-based Practice in social work is inconclusive. Reported results are mainly positive, however, only few studies apply a controlled design and most of the studies rely on self-report data. Furthermore, none of the studies assesses performance during real world practice, which is problematic considering that EBP as well as ESTs are skills that are supposed to be applied in practice. It is concluded that social work educators who want to base their teaching of EBP and/or the use of ESTs on research evidence should carefully assess the current body of research and draw on related research from other areas. Another conclusion is that more research with controlled designs and performance test measures, preferably in real-world settings, is needed.

For the *second study*, a case-based digital learning environment was developed with the goal to facilitate social work students' fallacy recognition skills. An experimental study with a 2x2 factorial design with two scaffolds, namely worked examples (with vs. without) and an external script (with vs. without), as between-subject factors was conducted with the goal to test the effectiveness of the scaffolds on students' ($N = 130$) fallacy recognition skills. Students analyzed vignettes of social work decision-making contexts. Each vignette entailed five different fallacies that were included in three different fallacious arguments from three different social workers. A fallacy was conceptualized as a violation of one of five criteria for good argumentation (Damer, 2011). Prior to the experiment, students watched a video with information on arguments and criteria for good argumentation. An additional baseline condition whose participants did not see the video was implemented. A post-test revealed that students from all conditions significantly increased their fallacy recognition skills on average. Students in the experimental conditions who saw the video outperformed students from the baseline condition who did not see the video. Therefore, Case-based Reasoning in conjunction with information about arguments and criteria for good argumentation (Damer, 2011) might be regarded as a promising instructional approach to facilitate social work students' fallacy recognition skills. However, the scaffolds did not contribute to further improvement of students' learning outcome. Hence, the study could not provide evidence regarding the effectiveness of worked examples or scripts on social work students' fallacy recognition skills beyond the use of Case-based Reasoning. This might be due to potential shortcomings in the designs of the worked examples and the script. For example, the worked examples were not very rich in that they did not show the solution path but only the solution. The script might not have taken learners' prerequisites sufficiently into account.

Within this doctoral thesis, the two empirical studies are framed by theoretical considerations (Part I) the studies are based on, as well as by a concluding section (Part III) in which a general discussion as well as respective implications for disciplinary and practical social work are provided.

1. Goals and Overview of the Dissertation

This chapter presents the problems that are addressed in this doctoral thesis as well as its goals that are derived from these problems. Subsequently, an overview of the doctoral thesis is provided.

Despite ongoing discussions within the social work discipline about social work's actual subject matter (e.g., Engelke, Borrmann, & Spatscheck, 2018; Lambers, 2018), one might hardly object the claim that one of social work's goals is to generate outcomes that social work clients value (Gambrill, 2013). It has been argued that both, the process of Evidence-based Practice (EBP; Gambrill, 1999) as well as Empirically Supported Treatments (ESTs; Macdonald, 1998) might contribute to good outcomes for social work clients by grounding social work practice in the best available (research) evidence. These assumptions are supported by some empirical evidence from fields other than social work such as medicine (e.g., Emparanza, Cabello, & Burls, 2015). Respectively – even though not without skeptical objections (e.g., Webb, 2001; see Gibbs & Gambrill, 2002 for a comprehensive summary of objections to EBP along with respective counterarguments) – there has been a considerable push in the social work literature to adopt related approaches (e.g., McNeece & Thyer, 2004) in social work practice, especially in the US (Ghanem, Lawson, Pankofer, Maragkos, & Kollar, 2017). However, empirical evidence suggests that the actual use in social work practice of both, EBP (e.g., Parrish & Rubin, 2012; Pope, Rollins, Chaumba, & Risler, 2011) as well as ESTs (e.g., Morago, 2010) lags. Accordingly, numerous empirical studies have been conducted to investigate potential barriers to the implementation of these approaches in social work practice (e.g., Aarons, 2004; Bellamy, Bledsoe, & Traube, 2006; Cawood, 2010; Gray et al.,

2012, 2015; Murphy & McDonald, 2004; Rubin & Parrish, 2010; Scurlock-Evans & Upton, 2015; Teater & Chonody, 2018; van der Zwet, Kolmer, & Schalk, 2016). Some of these barriers are insufficient preparation (Teater & Chonody, 2018), unsound training (Bellamy et al., 2006), negative attitudes towards EBP (Murphy & McDonald, 2004), diverse views of EBP (Rubin & Parrish, 2007) or lack of time for adequate preparation of ESTs (Cawood, 2010). It seems reasonable to assume that these barriers are, at least to some extent, related to insufficient or lack of education of EBP and/or ESTs in social work. However, it is yet unknown, how to best educate social work students and/or practitioners with regards to the acquisition of knowledge about EBP and/or the use of ESTs.

A second problem refers to the potential lack of/or potentially weak critical thinking (CT) skills of social work students, which are seen as prerequisites for the acquisition of EBP skills (Gambrill, 2019; Profetto-McGrath, 2005), for example in that they might help learners to value the benefits of EBP (Aglen, 2016). In a very narrow view, CT refers to the reduction of bias and the identification and explanation of fallacies (Gambrill & Gibbs, 2017). Respectively, educating social workers in skills related to the identification and explanation of fallacies has been promoted several times (e.g., Gambrill, 1994, 2012, 2019; Gambrill & Gibbs, 2017). Surprisingly, no empirical studies regarding education of these skills in social work were found in a literature review. However, studies from fields other than social work suggest that it is reasonable to assume that social workers need to be educated in identifying and explaining fallacies (e.g., Neuman, 2003; Neuman, Glassner, & Weinstock, 2004; Neuman, Weinstock, & Glasner, 2006; Neuman & Weizman, 2003). Yet, from an empirical perspective it is

unknown if social work students lack related skills and – if so – how to best teach them in these skills.

The first goal of the present doctoral thesis is to compile a systematic overview of empirical studies that investigate instructional approaches to foster knowledge about and/or skills in using EBP and/or ESTs in social work. The second goal is to systematically investigate these studies with respect to their characteristics as well as their quality to find out the best instructional approaches to teach the use of EBP and/or ESTs to social work students and/or practitioners. A third goal is to test the effectiveness of specific scaffolds (i.e., worked examples and external scripts) regarding social work students' skills related to the correct identification and explanation of fallacies in social work decision-making contexts within a case-based digital learning environment.

This doctoral thesis is separated in three parts. Part I investigates the literature relevant to the stated problems, thereby providing the theoretical foundation for two empirical studies, which are presented in Part II. Part III provides a comprehensive discussion on the results from both empirical studies as well as their respective implications.

Part I presents the theoretical background for the empirical studies (Part II).

Chapter 2 elaborates on the term *Evidence-based Practice* which, as a term, often is used to refer to two different approaches, *the process of Evidence-based Practice* as well as *Empirically Supported Treatments.* Both approaches are especially important for the first empirical study (chapter 6). Chapters 2.1 and 2.1.1 explain the history and main assumptions of the process of Evidence-based Practice. Since education of Evidence-

based Practice might contribute to its promotion and implementation, it is necessary to discuss whether Evidence-based Practice is a model that should be used in social work practice at all. Consequently, chapter 2.1.2 outlines criticism regarding the potential use of the process of Evidence-based Practice in social work contexts. Afterwards, chapter 2.1.3 provides empirical evidence regarding the current status of its use and implementation in social work practice. Subsequently, chapters 2.2 and 2.2.1 explain the roots as well as the main assumptions of Empirically Supported Treatments. Next, chapter 2.2.2 outlines critique regarding the use of Empirically Supported Treatments in social work practice. Afterwards, chapter 2.2.3 provides empirical evidence regarding the current use of Empirically Supported Treatments in social work practice. Finally, chapter 2.3 provides a summary of both approaches about the need for their implementation in social work practice, despite the criticisms offered.

Chapter 3 elaborates on the identification and explanation of fallacies in a social work decision-making context. These skills are related to argument evaluation and considered a prerequisite for the use of the process of Evidence-based Practice. Therefore, chapter 3.1 outlines what an argument is and how it is conceptualized within this thesis. Chapter 3.2 introduces the concepts of fallacies as well as five criteria for good argumentation, which are supposed to be beneficial for identifying and explaining fallacies. Chapter 3.3 provides a summary on argumentation and fallacies, which are especially important for the second empirical study (chapter 7).

Chapter 4 presents important theoretical accounts for the education of Evidence-based Practice and argument evaluation in social work. First, two instructional approaches are introduced in chapter 4.1, namely teacher-centered approaches (chapter

4.1.1) and student-centered approaches (chapter 4.1.2). Instructional approaches refer to the extent of freedom a learner is granted to structure her learning process. Student-centered approaches grant more freedom to the learners, while teacher-centered approaches structure the learners' learning processes to a higher degree. This differentiation is especially important for the first empirical study (chapter 6). Second, Case-based Reasoning is outlined as an example for a rather student-centered instructional approach that can be considered promising regarding the facilitation of social work students' fallacy recognition skills (chapter 4.2). The effectiveness of an instructional approach such as Case-based Reasoning might be enhanced by scaffolds. Consequentially, two scaffolds are accordingly explained: (1) Worked examples (chapter 4.3.1) and scripts (chapter 4.3.2).

Worked examples provide a learner with a problem, a solution process and an actual solution. Their conceptualization as well as their potential effects are explained in chapter 4.3.1.1. Effects of worked examples can be explained by Cognitive Load Theory which is outlined in chapter 4.3.1.2. While worked examples were initially primarily used in well-structured domains such as mathematics, chapter 4.3.1.3 illustrates more recent attempts of the successful application of worked examples in rather ill-structured domains such as medicine, to investigate whether they can be considered a promising approach for social work. Worked examples can be considered rather passive scaffolds, since they do not necessarily induce cognitive processes related to active problem-solving.

Therefore, *scripts* are then introduced as a more active scaffold (chapter 4.3.2). *External* scripts (i.e., scripts that are externally provided to a learner) provide knowledge to guide learners through cognitive operations. Their effectiveness is moderated by learners' *internal* scripts (i.e., knowledge structures that learners already have; chapter

4.3.2.1). The Script Theory of Guidance is outlined in chapter 4.3.2.2 as a theoretical basis that explains the interaction between internal and external scripts. Drawing on the Script Theory of Guidance as well as on empirical evidence, implications for the design of external scripts are provided (chapter 4.3.2.3).

Case-based Reasoning as well as scaffolds are especially important for the second empirical study (chapter 7).

Chapter 5 bridges chapters 1-4 with the empirical studies (chapters 6 and 7). It draws on the theoretical concepts provided in the first four chapters and outlines their implications for the empirical studies.

Part II entails two empirical studies that have been published in/or were submitted to scientific journals. Hence, each chapter contains a full-length scientific journal article.

Chapter 6 reports on the first empirical study which is a systematic review with the goal to shed light on the question *how to teach Evidence-based Practice in social work*. A systematic literature search for empirical studies was carried out and yielded 1116 hits. After double-coding the abstracts and subsequently 52 full articles, 28 articles (referring to 27 studies) met the eligibility criteria. The articles were double coded with reference to various dependent variables (e.g., effects of the conducted educational intervention on descriptive and procedural knowledge of Evidence-based Practice, the applied instructional approach, the methodological quality of the study). The results were evaluated against the following three research questions: (1) what educational interventions were applied and how were they conceptualized, especially with regard to

their instructional approach, (2) what were the effects of these interventions and (3) what was the methodological quality of the studies. Results are discussed with regards to promising instructional approaches for teaching Evidence-based Practice in social work. Gaps and needs for further research are derived from the discussion.

Chapter 7 presents the second empirical study, a randomized controlled trial with the objective to facilitate social work students' fallacy recognition skills. An experimental study with a 2x2 factorial design with an additional baseline group and the independent variables worked examples (with vs. without) and external script (with vs. without) was conducted in a digital case-based learning environment, with the goal to facilitate social work students' ($N = 130$) skills to identify and explain fallacies in social work decision-making contexts. The discussed results will hopefully help to develop effective instructional approaches to enhance social work students' fallacy recognition skills, which might eventually also contribute to the promotion of Evidence-based Practice in social work.

Part III discusses the results of the empirical studies (chapters 6 and 7) and provides implications for social work research and practice.

Chapter 8 provides a brief reminder on the dissertation's goals (chapter 8.1) and subsequently a discussion of the results (chapter 8.2) of the two empirical studies (chapters 6 and 7). Reasons are outlined that explain why the current body of research knowledge on education of Evidence-based Practice in social work can be considered rather weak. Furthermore, explanations for the (rather disappointing) effects of the worked examples and the external script are provided.

Chapter 9 is concerned with implications for social work research (chapter 9.1) and practice (chapter 9.2) that follow from the discussion of the studies' results (chapter 8.2). For example, it is argued that future research on education of Evidence-based Practice in social work is needed and that it should focus on the application of controlled designs to compare different instructional approaches while taking learners' prerequisites into account. Furthermore, it is argued that the second study (chapter 7) provides empirical evidence for the need to facilitate social work students' fallacy recognition skills. Thoughts are provided regarding the design of worked examples and external scripts for future research on that topic.

Chapter 10 offers some final conclusions by briefly summarizing and assessing the outcomes of the two empirical studies and emphasizing the importance for further research regarding education of Evidence-based Practice and its prerequisites (such as fallacy recognition skills) in social work as well as the potential contribution of this doctoral thesis in that regard.

2. Evidence-based Practice: Two alternative Conceptualizations

This chapter introduces Evidence-based Practice, which is a term that is often erroneously used for a variety of different concepts (e.g., Gambrill, 2006). However, two approaches seem to be the most prominent ones in social work: The process of Evidence-based Practice (EBP; Gambrill, 1999) and Empirically Supported Treatments (ESTs; Thyer & Myers, 2011). Many misconceptions in the literature regarding EBP have been identified (e.g., Gambrill, 2011) and different views of the model exist among practitioners (Rubin & Parrish, 2007). This confusion is also prevalent in disciplinary social work (Bates, 2006). For example, there are studies that survey social workers about their attitudes towards EBP (i.e., the investigated Evidence-based Practice approach is explicitly modeled as EBP in the report of the study), but instruments that were developed to investigate attitudes towards ESTs are applied (e.g., Booysen, Mbecke, Gouveia, Manomano, & Tanga, 2019). Likewise, there are reviews about Evidence-based Practice that conceptualize the model as EBP, but include also studies that investigated ESTs (e.g., Scurlock-Evans & Upton, 2015). Such confusion of the different approaches has also been observed in other disciplines such as psychology (Luebbe, Radcliffe, Callands, Green, & Thorn, 2007). It is thus necessary to accurately distinguish and describe these two approaches.

First, the origins (chapter 2.1) of EBP as well its main assumptions (chapter 2.1.1) will be explained, followed by critical considerations of the model (chapter 2.1.2) and empirical evidence regarding the status of the implementation of EBP in social work practice (chapter 2.1.3). Second, a brief historical sketch of ESTs will be provided (chapter 2.2). Afterwards, main assumptions of ESTs (chapter 2.2.1) and objections (chapter 2.2.2) as well as the status of their implementation in social work practice

(chapter 2.2.3) are outlined. Third, conclusions regarding the promotion of EBP and ESTs in social work are presented (chapter 2.3).

2.1 The Process of Evidence-based Practice

Evidence-based Practice (EBP) was initially developed in medicine. Guyatt (1991) coined the term Evidence-Based Medicine (EBM). Even though Guyatt coined the term, the model itself was built upon work from a group of scholars from the McMaster University (Evidence-based Medicine Working Group, 1992; Sackett & Rosenberg, 1995) and the history of the underlying paradigm can be traced back for centuries (see Claridge & Fabian, 2005 and Sur & Dahm, 2011 for overviews of the history of EBM). The founders of modern EBM – with Archie Cochrane and David Sackett being two key figures for the establishment of modern EBM (Claridge & Fabian, 2005) – intended to initiate a paradigm shift in the medical field in order to move from authority- and/or tradition-based decision-making and standard approaches to independent decisions that are based on critical appraisal of current research evidence, thereby providing superior patient care (Evidence-based Medicine Working Group, 1992). Subsequently, EBM sparked interest in the medical field and is nowadays considered a medical milestone (Dickersin, Straus, & Bero, 2007). The diffusion and increasing availability of the internet during the 1990s can be considered an important support for the potential of and increasing interest in EBM (Claridge & Fabian, 2005), since it allows practitioners to track down research evidence more efficiently (Gibbs, 2002).

Consequently, other professions became interested in EBM. By replacing EBM's target group – the *patients* – with a more neutral phrase such as *clients*, EBM became applicable for many human services and social care disciplines (Thyer & Myers, 2011).

Among them were nursing, psychology and public health (Satterfield et al., 2009). Furthermore, it was introduced to social work by Gambrill (1999). For all these other professions, the label of the model (i.e., Evidence-based Medicine) was changed to the more neutral term Evidence-based Practice, but the underlying model remained the same as in EBM. This model is explained in more detail below.

2.1.1 Main Assumptions from a Process Perspective on Evidence-based Practice

EBP describes a professional decision-making process that considers various sources of evidence by applying a specific sequence of cognitive operations. The process involves the "conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual [clients]" (Sackett, Richardson, Rosenberg, & Haynes, 1997, p. 2). With regards to evidence, original sources state that the consideration of research evidence, patients' preferences and clinical expertise is mandatory (Sackett et al., 1996). However, this model was updated by some of its originators by replacing clinical expertise with clients' circumstances, thereby giving greater emphasis on the clients' perspective (Haynes, Devereaux, & Guyatt, 2002).

Consequently, c*linical expertise* (i.e., basic skills of clinical practice, including relationship skills and the practitioner's individual experience; Haynes et al., 2002) was now no longer a source of evidence, but a means for integrating three different kinds of evidence, namely (1) research evidence, (2) clients' circumstances and (3) clients' preferences into an informed and mutual (i.e., together with the client) decision-making process.

Regarding *research evidence*, attempts have been made to formulate hierarchies of the quality of such knowledge. The logic of such hierarchies circles around the potential internal validity of different research approaches. Thus, systematic reviews/meta-

analyses are likely to be considered at the top of such hierarchies with randomized controlled trials as second best evidence (e.g., McNeece & Thyer, 2004). In contrast, at the bottom of such hierarchies are usually approaches such as qualitative studies (McNeece & Thyer, 2004) or expert opinion (Straus, Richardson, Glasziou, & Haynes, 2005), depending for example on the field or authorship. While EBP refers to the best *currently available* evidence, such lower located sources might very well provide important evidence. However, what counts as *best* evidence depends eventually on the question at hand (Gambrill, 2013). Questions in social work may, amongst others, refer to effectiveness, prevention, assessment, cost, practice guidelines or self-development. Evidence hierarchies typically refer to the selection of an intervention for an individual and may, for example, not be appropriate for the selection of policies (Gambrill, 2013). Therefore, the question what counts as best evidence is (at least in social work) context-bound and should be answered with respect to each individual case. Additionally, not only evidence but also uncertainty and ignorance (e.g., informing a client that a search was conducted and revealed that no research evidence related to her problem is currently available; Gambrill, 2013, 2014) are shared with the client.

Clients' circumstances refer to questions such as where a client lives (e.g., a rural area) and potential consequences such as barriers regarding the continuous access to a specific intervention (Haynes et al., 2002).

Clients' preferences refer to "the unique preferences, concerns and expectations each [client] brings to … an encounter and which must be integrated into … decisions if they are to serve the [client]" (Straus et al., 2005, p. 1).

As stated before, EBP describes a process of cognitive operations. This process includes five steps as described in original sources (Sackett et al., 1997):

1. Converting information needs related to practice decisions into well-structured questions.

2. Tracking down, with maximum efficiency, the best evidence with which to answer those questions.

3. Critically appraising that evidence for its validity, impact (size of effect), and applicability (usefulness in practice).

4. Integrating this critical appraisal with clinical expertise and with a client's unique circumstances and characteristics including their values and preferences and making a decision together with the client.

5. Evaluating the effectiveness and efficiency in carrying out Steps 1-4 and seeking ways to improve them in the future.

At present it is mainly a question of plausibility when it comes to deciding whether one should implement and promote EBP in social work practice settings or not, since empirical evidence regarding the effectiveness of EBP is largely missing. Consequently, the introduction of EBP to social work (Gambrill, 1999) triggered vital discussions among social work scholars on whether the social work discipline (and eventually social work practice) should embrace the model or not. Ever since EBP received a considerable amount of criticism which will be outlined now.

2.1.2 Criticism of the Process of Evidence-based Practice

Criticism regarding EBP can roughly be categorized in (1) philosophical (i.e., epistemic) considerations, (2) barriers to the implementation of EBP and (3) misconceptions about EBP.

A first concern about EBP that is related to *philosophical considerations* is the assumption that EBP implies a positivistic paradigm which leads to the inappropriate reduction of social processes in order to make them measurable. Accordingly, some scholars suppose that EBP might be insufficient to investigate the full complexity of social problems (Albus & Micheel, 2012; Gitterman & Knight, 2013; Webb, 2001), due to its rather atheoretical nature and strong empirical focus (Adams, Matto, & LeCroy, 2009; Gray & McDonald, 2006; Green, 2000). For example, Adams et al. (2009) are concerned that valuable theoretical knowledge might not be found in searches for empirical evidence. This criticism is related to the proposed high complexity of social problems (e.g., Staub-Bernasconi, 2018). Some scholars argue that social problems imply numerous causal relationships due to their complexity. Therefore, it is argued that a simple input-output approach is hardly applicable in order to evaluate social work interventions (Gray & McDonald, 2006). Regarding EBP, this might be problematic in cases where the search for the best currently available research evidence suggests a respective social work intervention as the potentially best treatment for a client, and the intervention is implemented without further theoretical considerations. Thus, Green (2000) outlines that theories are needed to investigate whether a given program is adequate, despite the question of mere effectiveness. For example, a program which is intended to promote condom use might be effective, but further theoretical considerations must be taken into account (e.g., questions such as *do service recipients have the skills to negotiate condom use with their partner*). However, it could be argued that EBP fulfills this demand by explicitly taking clients' circumstances into account.

A second criticism relates to EBP's origin in medicine. In this regard, Adams et al. (2009) state that it is a medical model that follows a respective focus on individuals, which is considered as insufficient for social work, because there, practitioners often

work with client systems (e.g., a family) instead of a single client. While it is essentially true that social workers often work with client systems, this argument provides no further reason that explains why the model should not be applicable in such contexts.

A third source for criticism refers to the paucity of EBP. Indeed, there is some slight empirical evidence for EBP being superior compared to a non-EBP approach regarding the generation of favorable outcomes for patients in the field of medicine (Emparanza et al., 2015). However, apart from that, the assumption that EBP is more beneficial than other approaches is mainly based on opinion (Thyer, 2015) and plausibility (Hüttemann, 2006). This circumstance was also addressed by the founders of EBP in one of their first articles on the approach, in which they state that it will be merely possible to investigate the effectiveness of EBP empirically (Evidence-based Medicine Working Group, 1992).

A fourth concern is related to the question what counts as *currently best available evidence*. For example, several scholars have raised concerns about the external validity of randomized controlled trials with regards to social work contexts (e.g., Cartwright & Munro, 2010; Mullen, 2016; Otto & Ziegler, 2008). If, for example, a randomized controlled trial provides strong evidence for the efficacy of a certain intervention for a sample drawn from US citizens, this must not necessarily be assignable to, for example, a polish population. However, promoters of EBP such as Thyer (2001) state that in such a case, the evidence from the US sample would simply count as the currently best available evidence, and that this is better than no evidence at all. Otto and Ziegler (2008, p. 274) reject the idea of currently best available *evidence* to some extent, when they state that "evidence about what works based on poorly designed studies and invalid, biased data is worse than no research evidence at all." This remains an open and rather philosophical question, but it seems reasonable to prefer a well-elaborated and plausible theory as a

source of "evidence", in cases where empirical evidence is only available from poorly designed studies.

Barriers to the implementation of EBP can be split into two different categories: (a) Barriers to the implementation of the general EBP process and (b) barriers to the application of evidence found through the EBP process.

Empirical evidence, mostly coming from surveys among practitioners, suggests that *barriers regarding the implementation of the general process of EBP* are for example practitioners' lack of time to engage in EBP in practice, limited amount of relevant research evidence, lack of organizational support, negative attitudes towards EBP as well as lack of knowledge, skill and training (see Gray et al., 2015; Mullen & Streiner, 2004; Scurlock-Evans & Upton, 2015; Straus & McAlister, 2000; Wike et al., 2014 for comprehensive reviews regarding EBP barriers). Furthermore, Albus and Micheel (2012) point to the problem of resources needed for research on effective programs, practice evaluation and controlling. Accordingly, these authors argue that EBP is simply too cost-intensive to be implemented properly. This is a valid argument, however, improving outcomes that are valued by clients will probably always be associated with investment. Therefore, the general argument of increased costs is certainly not exclusive to EBP, but most likely applicable to any attempt regarding the improvement of the quality of social work's outcomes. It is nevertheless an open debate whether instruments such as research on the effectiveness of programs, practice evaluation or controlling are generally reasonable in social work or not (e.g., Otto, Polutta, & Ziegler, 2010). However, it could be argued that these instruments refer more to ESTs and can thus – to some extent – also be considered as misconceptions with regard to a discussion that focuses on EBP.

Regarding *barriers to the application of evidence found through the EBP process*, founders of EBP conceptualized the four so-called "killer B's" (Sackett, Straus, Richardson, Rosenberg, & Haynes R. B., 2000, p. 179; "B" refers to the picture of a bee potentially buzzing in one's ear when considering the application of evidence in a given situation; the respective buzzing is meant to be a sign for increased caution). These four killer B's are: (a) *burden* of illness (i.e., the frequency of a disorder in a given area is too low to warrant implementation or the hoped-for-outcome is unlikely to occur in our client), (b) the *beliefs* of the client are incompatible with the application of the evidence, (c) implementation costs constitute a bad *bargain* for individual and/or community's resources and (d) other geographical, organizational, traditional, authoritarian, legal of behavioral *barriers* are so high that it is not worth trying to overcome them.

Some aspects of criticism about EBP have been labeled as *misconceptions*, since they stem arguably from confusion about the different approaches of Evidence-based Practice. One reason for such misconceptions might be inaccurate descriptions in scholarly literature. For example, whole books about EBP such as *Evidence-based social work: A critical stance* by Gray, Plath, and Webb (2009) do not describe the five steps of EBP, which are a core concept of the approach (cf., Gambrill, 2010).

A very common misconception refers to the confusion of central properties of EBP with those of ESTs. For example, practitioners are often unsure whether EBP refers to a decision-making process (EBP) or to interventions someone deemed as effective based on research evidence (ESTs; e.g., Rubin & Parrish, 2007; Thyer, 2015; van der Zwet, Kolmer, Schalk, & van Regenmortel, 2019). Other examples for misconceptions are such claims as that EBP denigrates expertise, that it ignores clients' values and preferences, that it is simply a cost-cutting tool, that it is an ivory tower concept that cannot be done,

that it is limited to clinical research and that it leads to therapeutic nihilism if no research evidence for the problem at hand is available (Mullen & Streiner, 2004; Straus & McAlister, 2000).

Another misconception is that only evidence from randomized controlled trials (or respectively reviews that summarize such studies) is considered as "good" evidence within the EBP framework (Gibbs & Gambrill, 2002). This argument might be related to the confusion of EBP and ESTs, since randomized controlled trials were promoted as the gold standard for the scientific investigation of the effectiveness of treatments (see chapter 2.2.2). However, it is more likely that it stems from evidence hierarchies (see chapter 2.1.1) that were introduced in the medical field regarding EBP (Straus et al., 2005). Such hierarchies might seem reasonable in medicine, where efficacy of a treatment (e.g., a drug) for a given population of patients is of utmost interest and randomized controlled trials might be a methodologically appropriate instrument to gather respective evidence. However, this might not be necessarily true for social phenomena (e.g., social problems) that tend to be much more complex. Nevertheless, similar hierarchies of evidence have been promoted for EBP in social work too (e.g., McNeece & Thyer, 2004), which is perceived as inadequate by many social work scholars (e.g., Gambrill, 2013; Webb, 2001). It is true that EBP favors methods that can test claims in a rigorous way, but the problem at hand is crucial for answering the question what counts as best evidence (Gambrill, 2013; Gibbs & Gambrill, 2002). In social work, such problems can, as stated earlier, refer to vastly different questions. Therefore, nowadays even strong proponents of EBP doubt that such evidence hierarchies are helpful in social work contexts (e.g., Gambrill, 2013).

Finally, many scholars state that EBP negates the value of theories due to its focus on empirical evidence (e.g., Green, 2000). In social work, theories are often claimed to be

of significant importance (e.g., Munro, 2002). Therefore, it is not surprising that EBP is recognized to a lesser extent and discussed with even greater skepticism in countries with a strong theory-based tradition in social work discipline such as German-speaking countries, as compared to countries with a traditionally greater emphasis on empirical research such as the US (Ghanem et al., 2017). However, while prominent proponents of EBP identify EBP itself as *atheoretical*, they acknowledge at the same time, that EBP does not exclude the consideration of relevant theory (Thyer, 2013). This is especially important, since it might be the case that searches for empirical evidence regarding questions related to social problems might yield scarce or even no results (Gibbs & Gambrill, 2002). Nevertheless, ethical codes such as the National Association of Social Workers (2017) from the USA or the Deutscher Berufsverband Sozialer Arbeit e.V. (2014) from Germany include the statement that social workers are supposed to consider research evidence to guide their practice. In general, this requires social workers to search for relevant research findings and share the results with their clients (cf., Gambrill, 2010). In some cases, this might require the social worker to share the information that a search has been conducted, but no results were found (Gambrill, 2014; Gibbs & Gambrill, 2002).

Taken together, EBP is confronted with a lot of criticism. Some of it is indeed legitimate and offers food for thought regarding further development of the approach and important aspects to consider with its implementation in social work practice. However, much of it relates to misconceptions. Furthermore, it seems reasonable to argue for an ethical obligation of practitioners in a helping profession such as social work to consider the best currently available research evidence when engaging in mutual decision-making processes with clients. Therefore, the implementation of EBP seems not only to be a

promising endeavor regarding the generation of outcomes that are valued by clients, but also with respect to the ethical code of the social work profession.

It is thus an interesting question to which extent EBP is used in social work practice. Empirical evidence regarding the current status of EBP implementation is presented in the next section.

2.1.3 Current Status of the Implementation of the Process of Evidence-based Practice in Social Work Practice

Much of the empirical evidence regarding the implementation of EBP in social work practice comes from the United States, where EBP has been discussed earlier and more extensively than, for example, in German-speaking countries (Ghanem et al., 2017). Empirical evidence suggests that scholars in Germany still seem to discuss whether EBP can be considered a promising and valuable approach, whereas it seems that scholars in the US have overcome this discussion and are currently thinking about ways to promote its implementation (Ghanem et al., 2017). Mullen and Streiner (2004) argue that, while EBP appears to be mostly prominent in the US, Canada and England, countries such as Sweden, Finland, Norway and Denmark show growing interest. These differences with respect to countries might relate to different cultures in social work and it is not really surprising to see such discrepancies, considering that there are ample differences in social welfare even between countries that are geographically close to each other such as Canada and the US (Holosko & Leslie, 2001).

Regarding the *implementation of EBP in the United States*, Pope et al. (2011) found in a survey of social workers (N = 200) in the United States that, although 83% agreed or somewhat agreed to be familiar with social work databases, only 56% agreed

or somewhat agreed that they used relevant research to answer clinical questions (range: agreed, somewhat agreed, undecided, somewhat disagree, disagree). Knight (2013) conducted a survey with 151 social workers from the US. Half of the participants felt that their social work education did not prepare them to read research articles and over 90% "agreed" or "strongly agreed" (range: strongly disagree, disagree, agree, strongly agree) that they avoid to read articles that report on empirical research. A similar survey among social work students ($N = 421$) who were engaged in field practice yielded comparable results (Knight, 2015). Half of the students felt "not at all" or "not much" prepared to engage in EBP. While more than 80% "agreed" or "strongly agreed" that they engage in EBP, only few reported to engage in related practice behaviors (e.g., use research literature to guide own practice, evaluate effectiveness of their own practice). In a survey by Parrish and Rubin (2012) of 688 social workers carried out in Texas, it was found that few social workers indicated on a 5-point Likert-type scale that they "often" or "very often" use the Internet to search for the best evidence to "guide practice decisions" (32.8%), "read about research evidence to guide practice decisions" (37.8%), "inform clients of the degree of research evidence supporting alternative intervention options" (25.6%), and "engage in all steps of the EBP process" (15.1%).

As for empirical evidence regarding the *implementation of EBP in countries other than the United States*, Heiwe et al. (2013) conducted a survey among 174 medical social workers in Sweden. 75.3 % of the participants reported a need to increase the use of evidence in their daily practice. In a survey among Norwegian social workers ($N = 2033$) by Ekeland, Bergem, and Myklebust (2019), 64.8% reported to use research literature only "a couple of times per year" or "almost never". Likewise, van der Zwet et al. (2016) conducted a survey among social workers ($N = 341$) in the Netherlands. They found that,

even though social workers' familiarity with and attitudes towards EBP were slightly positive, their intentions to engage in EBP as well as their actual engagement were relatively low. Interestingly, a comparison of these data with survey data from Master of Social Work (MSW) students ($N = 32$) revealed that MSW students show significantly higher intentions to engage in EBP as well as actual engagement (van der Zwet, Weling, Kolmer, & Schalk, 2017). Thus, it could be speculated that the next generation of social workers is more eager to include research evidence in their daily practice as compared to older generations. Correspondingly to van der Zwet et al. (2016), van der Zwet et al. (2019) found in a series of interviews with social workers ($n = 12$) and other staff from social services ($n = 10$) that the majority (19 out of 22) thought it was important that their organization would engage more in EBP.

Taken together, it is unclear to what extent social work students and practitioners use EBP in practice (or if they do at all). However, the available empirical evidence suggests that the current state of implementation of EBP in social work practice can be considered as relatively weak. At the same time, some findings suggest that the next generation of social workers (i.e., students) has favorable attitudes towards and intentions to use EBP. Furthermore, barriers to EBP implementation such as reported lack of knowledge, skill and training or misconceptions in general underpin the necessity of effective EBP education in social work.

2.2 Empirically Supported Treatments

The American Psychology Association consists of more than 50 so-called Divisions. During the middle of the 1990s, the president of Section III (Society for a Science of Clinical Psychology) of Division 12 (Clinical Psychology) assembled a Task Force. Its goal was to promote and disseminate psychological procedures by developing standards of evidence, which would give these procedures a status of being *empirically validated* (later changed to the more tempered term *empirically supported*; Chambless, 1996; Garfield, 1996) as well as by preparing a list of such procedures that would meet these very standards (Thyer & Myers, 2011). These standards are explained now.

2.2.1 Main Assumptions from an Interventional Perspective on Evidence-based Practice

The Task Force (1993, p. 10) established two different criteria of evidence, one for (1) Well-Established Treatments and one for (2) Probably Efficacious Treatments.

1. *Well-Established Treatments* fulfill the following criteria:

    I. At least two good group design studies, conducted by different investigators, demonstrating efficacy in one or more of the following ways:

        A. Superior to pill or psychological placebo or to another treatment.

        B. Equivalent to an already established treatment in studies with adequate statistical power, or

    II. A large series of single case design studies demonstrating efficacy. These studies must have:

        A. Used good experimental designs and

        B. Compared the intervention to another treatment as in I.A.

    III. Studies must be conducted with treatment manuals.

IV. Characteristics of the client samples must be clearly specified.

2. *Probably Efficacious Treatments* have to fulfill less stringent criteria:

I. Two studies showing the treatment is more effective than a waiting-list control group, or

II. Two studies otherwise meeting the well-established treatment criteria I, III, and IV, but both are conducted by the same investigator. Or one good study demonstrating effectiveness by these same criteria, or

III. At least two good studies demonstrating effectiveness but flawed by heterogeneity of the client samples, or

IV. A small series of single case design studies otherwise meeting the well-established treatment criteria II, III, and IV.

Empirically supported treatments (EST) sparked interest, as indicated by entire issues of important journals such as *Clinical Psychology: Science and Practice* in 1996, the *Journal of Consulting and Clinical Psychology* in 1998 or the Journal *Psychotherapy Research* in 1998. Ever since, the list of such treatments was consequently updated (e.g., Chambless et al., 1996; Chambless et al., 1998) and at the time of writing this dissertation, the current list of now so-called Research-Supported Psychological Treatments is accessible via the internet page of Division 12 (https://www.div12.org/psychological-treatments/).

Subsequently, the concept of ESTs was recognized and – to some extent – also embraced by other fields. For example, the federal Department of Education in the US supports a *What Works Clearinghouse* (https://ies.ed.gov/ncee/wwc/). This institution provides an overview of many ESTs that are supposed to be effective in education, for

example in order to deal with problematic classroom behavior, drop-out prevention or teaching mathematics. Likewise, and related to social work, the *California Evidence-based Clearinghouse for Child Welfare* (https://www.cebc4cw.org/) lists many different ESTs relevant to social problems in child welfare. Both websites also allow the user to compare the ESTs against each other with regards to their relative effectiveness.

However, in contrast to these favorable reactions, the effort to establish standards of evidence for ESTs sparked also a considerable amount of controversy from researchers as well as from practitioners (Herbert, 2003) which is outlined now.

2.2.2 Criticism of Empirically Supported Treatments

After publishing their evidence standards for ESTs, the Division 12 was confronted with severe criticism (Garfield, 1996). This criticism referred mainly to four different aspects: (1) The overall EST concept itself, (2) the standards for evidence, (3) the requirement of treatment manuals and (4) the promotion of randomized controlled trials as the gold standard to evaluate treatments.

Issues that referred to the *overall concept of ESTs* were, for example, problematic language (since *validated* seems to imply proven and guaranteed effectiveness), the fact that patients are different and that treatments usually produce different results, problems regarding reliability and validity of studies that validate treatments, lack of differential effectiveness (i.e., different treatments tend to be equally effective, maybe due to the important contribution that the relationship between the therapist and the patient might have on the desired outcomes) or unknown variables that also contribute to effectiveness (Garfield, 1996).

Further aspects of criticism concerned the *standards of evidence*, such as that the sufficiency of only two supportive studies promote bias against the publication of null results (since other studies might not have detected any effects) and that the allowance of passive control groups (for the label *probably efficacious*) is problematic, since almost any treatment in psychotherapy is superior to no treatment (Castelnuovo, Faccio, Molinari, Nardone, & Salvini, 2004; Herbert, 2003). It was furthermore criticized that the EST approach promotes quick symptom reduction over the long-term establishment of quality of life (Castelnuovo et al., 2004).

A target of particular criticism were the *treatment manuals*, which were deemed very important by the Division 12, since they are supposed to contribute to fidelity as well as to further investigation of the treatments by clearly describing what was actually done during the treatment (Chambless et al., 1996; Chambless & Hollon, 1998). Garfield (1996) questioned the external validity of these manuals, especially since practitioners are usually not sufficiently trained in respective treatments. Other criticism was that manuals value fidelity more than competence and that they promote *schoolism* in psychotherapy.

The Division 12 rejected most of these criticisms and argued that much of it was based on misconceptions (Chambless, 1996). However, the unfortunate choice of the term *validated* was acknowledged. They assured that they did not intend to limit practice to ESTs, stated that manuals can only be effective if critical components are known and that flexibility of the therapist in charge is important if unexpected events occur. Nevertheless, they emphasized their belief that both, relationship and treatment, contribute to outcome (Chambless, 1996). They also mentioned, that the first step is to

focus on *efficacy* (i.e., how a treatment performs in a controlled environment such as a randomized controlled trial), before investigating *effectiveness* (i.e., the extent to which treatments are effectively transferable into real practice scenarios; Chambless et al., 1998). Consequently, the list of ESTs was updated (Chambless et al., 1996; Chambless et al., 1998; Chambless & Ollendick, 2001) and suggestions for efficacy research were made (e.g., importance of replication, randomized controlled trial as a preferable design, high power; Chambless & Hollon, 1998; Chambless & Ollendick, 2001).

The subsequent *focus on randomized controlled trials* as the gold standard for the evaluation of treatments also received severe criticism. Critics claimed that randomized controlled trials are not capable of capturing the factors that are causal for change within complex therapist-client-relationships (Bohart, O'Hara, & Leitner, 1998), that more complex problems and interventions that are more difficult to measure are excluded (Otto et al., 2010), that some important features of randomized controlled trials such as randomization or standardization of treatment procedures are hard to come by in psychotherapy studies (Castelnuovo et al., 2004; Castelnuovo, 2010), or that randomized controlled trials lack external validity (Cartwright & Munro, 2010).

Some of these arguments are supported by empirical evidence. One example relates to the reliability of treatment manuals. Ablon and Jones (2002) conducted a study in which expert therapists developed prototypes of brief interpersonal psychotherapy as well as cognitive behavior therapy. The prototype transcripts were then compared to the actual therapy administration by external experts. Both prototypes adhered (during their administration) most strongly to the prototype of cognitive behavior therapy. They also shared many interactional characteristics such as offering advice and guidance or

promoting change in behavior. Especially such characteristics, essentially supportive in nature, were correlated with positive outcomes. Ablon and Jones (2002) assume that one possible explanation for the vast amount of similarities among two – at least on the surface – very different approaches could be that proponents of different forms of psychology might apply different terminologies for rather similar constructs. Such observations cast doubts on the reliability of treatment manuals. Consequently, it is questionable if randomized controlled trials are an adequate means to investigate psychological treatments, when a key feature of randomized controlled trials – the standardization of the treatment – is not fulfilled.

Much of this criticism, initially formulated in psychology, has been brought up in social work as well (e.g., Borrmann & Thiessen, 2016; Otto et al., 2010; Otto & Ziegler, 2008; Sommerfeld & Hüttemann, 2007). Especially the criticism regarding randomized controlled trials seems reasonable with respect to complex fields such as psychotherapy or social work. Randomized controlled trials might, for example, be ideal in medicine to test the efficacy (and, in medicine, thereby presumably also the effectiveness, at least to some extent) of drugs. But to maintain external validity in messy fields such as social work is particularly more difficult. While the Division 12 acknowledged that testing the efficacy of treatments can only be a first step, the question of their effectiveness has not been sufficiently addressed yet.

Taken together, much of this critique on ESTs seems to be valid and points to the necessity of further development of means to evaluate the effectiveness of ESTs. Randomized controlled trials with small samples, poorly designed studies or client and

therapist variables are potential threats to the validity of such evaluations and might yield inflated claims regarding the effectiveness of ESTs.

However, it is worth noting that some treatments were also demonstrated to be ineffective or even harmful in the past (see Barlow, 2010 for a historical perspective on related psychological research and Lilienfeld, 2007 for a review and list of potentially harmful therapies). Furthermore, failure to check effects of interventions have led to much harm, such as blinding 10.000 babies by giving them oxygen at birth (Silverman, 1980) and many other attempts that suggest that good intentions do not necessarily lead to good outcomes (see Gibbs, 2002 for examples throughout the course of history). Therefore, it seems plausible to try to enhance research methods and designs to fit them more properly to the needs that are required to investigate treatments thoroughly, thereby providing strong evidence that empirically supports (or impairs) such treatments. Hence, a thoughtful promotion of ESTs that takes their current limitations into account appears to be a beneficial addition to the field of social work.

Likewise, leading social work scholars argue that social work clients have a right to receive effective treatment (Gambrill, 2013; Thyer, 2015). At the same time, these scholars admit that carrying out well-conducted randomized controlled trials in ill-structured domains is challenging, but not impossible (Gambrill, 2013). The call for the use of ESTs raises the question, to which extent social workers use ESTs in social work practice. Empirical evidence regarding this question will be briefly presented and discussed now.

2.2.2 Current Status of the Use of Empirically Supported Treatments in Social Work Practice

Claims have been made that suggest a great impact of ESTs, at least in the US (James, Lampe, Behnken, & Schulz, 2019; Thyer & Myers, 2011). However, respective empirical evidence is scarce and inconsistent.

With regards to the *United States*, Pignotti and Thyer (2012) conducted a survey among social workers ($N = 400$) from 39 different states in the US. The vast majority (97.5%) reported that they use at least one EST. However, also 86% reported to use at least one treatment that lacks sufficient supportive empirical evidence. Among these were also treatments that are considered potentially harmful. One example for such a treatment is Critical Incident Stress Debriefing (Lilienfeld, 2007) that was used by almost 25% of the respondents. Empirical evidence that contradicts the findings from the survey by Pignotti and Thyer (2012) to some extent comes from Horwitz et al. (2014). They conducted interviews with 184 informants (mostly directors) from 83 public child welfare agencies regarding the implementation of ESTs. 94% of the agencies reported that they had started a new program within the last five years, however, only 25% of these programs were evidence-based.

Empirical evidence regarding the use of ESTs in countries *other than the United States* is rare. Morago (2010) reported that 42.6% of 155 social workers and social care professionals indicated the level of implementation of ESTs in their respective agency as "very poor" and 40% as "modest" in a survey conducted in the UK (range: very poor, modest, good, excellent). James et al. (2019) conducted a survey in Germany with the objective to investigate participants attitudes towards ESTs (the term was changed to *research -based practice methods* due to the assumption that social workers in Germany

were not familiar with the original terminology). Only about 40% indicated a great or very great (range 0 = "not at all" to 4 = "very great") openness to try research-based practice methods.

The overwhelmingly high number of respondents who indicated that they use at least one EST in the survey by Pignotti and Thyer (2012) raises the question how social workers are trained in ESTs. The gold standard for training of new treatments is a combination of a didactic program and supervised clinical work (Weissman et al., 2006). Interestingly, Weissman et al. (2006) conducted a survey among training directors of 64 MSW programs in the US and found that 62% did not require this gold standard. The authors concluded that the research-practice-gap will remain without a considerable increase of EST-training. Cawood (2010) conducted a survey among 250 school social workers regarding potential barriers to the use of ESTs. 70% think that lack of time for adequate preparation as a respective barrier and 54% lack of knowledge regarding ESTs.

Taken together, the limited empirical evidence about the implementation of ESTs in social work suggests that there is a certain push regarding EST implementation, at least in the US. At the same time, some studies point out a lack of appropriate education regarding the use of ESTs in social work. This points towards the need for effective education in the use of ESTs.

2.3 Summary

Chapter 2 introduced two important concepts of this dissertation, the process of Evidence-based Practice (EBP; chapter 2.1) and Empirically Supported Treatments (ESTs; chapter 2.2). Both are considered promising means to decrease the research-

practice-gap, despite a considerable amount of criticism that has been raised against both approaches. Chapter 2.1.1 outlined EBP as a decision-making process that entails five unique steps with the goal to integrate the currently best available research evidence with clients' circumstances and preferences based on clinical expertise. A wide array of objections to EBP has been explained (chapter 2.1.2). These objections refer either to philosophical implications of the concept, barriers regarding its practical applicability or misconceptions. Nevertheless, it was concluded that the implementation of EBP is a promising endeavor that deserves to be supported. Empirical evidence was outlined to investigate the current status of the implementation of EBP in social work practice (chapter 2.1.3). Respective evidence suggests that EBP is barely used in social work practice and that a part of this problem is related to the lack of/ or suboptimal EBP education. Chapter 2.2 provided a short history of the establishment of ESTs. ESTs were introduced as treatments that fulfill one of several sets of criteria regarding empirical evidence related to their efficacy (chapter 2.2.1). Different sets of criteria refer to different levels of efficacy. Criticism of ESTs refers especially to their standards of evidence. For example, many scholars doubt that randomized controlled trials, which are considered the gold standard for their evaluation, are capable of accurately investigating their actual effectiveness (chapter 2.2.2). It was argued that a thoughtful promotion of ESTs is recommended for social work practice. Subsequently, empirical evidence regarding their current use in social work practice has been analyzed (chapter 2.2.3). This evidence suggests that the use of ESTs is higher in the US than in other countries (e.g., Germany). However, empirical evidence suggests also a wide use of treatments that lack any empirical foundation regarding their efficacy or effectiveness and some of them can be considered potentially harmful by empirical standards. Suboptimal education of

ESTs was identified as one potential component for the lag of EST implementation in social work.

It was proposed that the helping professions have an ethical obligation to base their practices and methods on the best evidence that is currently available. A lack of sufficient education has been identified as one key factor contributing to the insufficient use – or arguably general absence – of EBP and/or the use of ESTs in social work practice. From this perspective, it is important to promote and facilitate the implementation and appropriate use of both, EBP as well as ESTs in social work. The effective education of EBP and ESTs is one important aspect in that regard.

3. Argument Evaluation as an Important Prerequisite for Evidence-based Practice

One important goal of EBP in social work is to arrive at evidence-informed decisions related to social problems of clients in order to achieve valued outcomes (Gambrill, 2019). The process to arrive at such decisions involves reasoning as well as its externalization (i.e., argumentation), since they are usually a mutual product of at least one social worker and a client (and often involve other stakeholders such as other professionals or members of the client system). However, arguments are prone to numerous fallacies (Damer, 2011; Walton, 1995). A fallacy can be seen as a mistake in argumentation that does not necessarily appear to be a mistake (Damer, 2011), or as an argumentative move with the intention to block critical appraisal and to get the best out of one's own argument (Walton, 1995).

Skills related to evaluate arguments (e.g., identify and explain fallacies) are closely related to critical thinking (Gambrill, 2019; Gambrill & Gibbs, 2017). Critical thinking is considered an important prerequisite for EBP (Aglen, 2016; Gambrill, 2019; Profetto-McGrath, 2005). Hence, fostering skills related to the identification and explanation of fallacies in social work decision-making contexts are one important first step regarding the facilitation of EBP.

Chapter 3 introduces theoretical foundations with respect to argumentation and fallacies. First, it will be outlined how to build arguments, how they differ in terms of their underlying structure (deductive versus inductive) and what tasks are involved through the course of argumentation (chapter 3.1). Second, the concept of fallacies is defined as a violation of one of five criteria for good argumentation (chapter 3.2). Third, criteria for good argumentation as well as their underlying principles are explained (chapter 3.2).

3.1 The Importance of Argumentation in EBP

The process of Evidence-based Practice (EBP) defines a decision-making process with the goal to solve clients' problems (e.g., Gambrill, 2019; Sackett et al., 1996; Sackett et al., 1997; Straus, Richardson, Glasziou, & Haynes, 2011). In social work, problem-solving and decision-making are closely related to argumentation, since they are processes that include the exchange of arguments between social workers, clients and other stakeholders (Gambrill, 2019). For example, EBP values clients' unique characteristics, such as their preferences and circumstances. Thus, integrating these with the best currently available research evidence based on the social worker's expertise (step four of the EBP process) involves necessarily the exchange of arguments, for example, between a social worker and a client or between two or more social workers, since often, more than one social worker is involved in professional decision-making processes. To carry out argument evaluation appropriately, it is helpful to know (1) how to build arguments, (2) how they can be classified and (3) what tasks are included in argument evaluation.

First, *an argument consists of* a conclusion and one or more claim/s that is/are supposed to support the conclusion (Damer, 2011). The goal of argumentation is to convince other parties of one's own propositions based on their claims (Toulmin, 2003). Claims can be either explicit or implicit (i.e., not explicitly stated) and furthermore be conclusions themselves (Damer, 2011). Consider the following example:

> *"It was only three months after Harold got married that he started smoking pot. His wife must have gotten him started on the stuff."*

This argument can be broken down into standard form in the following way:

"Since Harold recently got married, (explicit claim)
and having a new wife preceded the event of starting to smoke pot, (explicit claim)
[and what precedes another event is the cause of that event,] (implicit claim)
Therefore, Harold's wife is the cause of his pot smoking." (explicit conclusion)

The implicit claim is not stated in the original argument, but it is implied to bridge the gap between the stated claims and the conclusion. Furthermore, the premise *what precedes another event is the cause of that event* is not only a claim, but also a conclusion itself. However, it is used as a claim to support the overall conclusion of the argument (i.e., that Harold's wife is the cause of his pot smoking).

The premise *what precedes another event is the cause of that event* is, of course, not always necessarily true. In case the claim is indeed false, the very argument can be considered *fallacious*. Considering that this is the case in the former example, a social worker who does not recognize the fallacy (i.e., the fallacious argument) might conclude that it is justified to advise Harold to quit the relationship with his wife in order to quit pot smoking (this example might seem very unlikely, however, one might think of a client in a probation period in which pot smoking might have severe judicial consequences).

Second, *arguments can be broadly classified* in two ways: (a) Deductive and (b) inductive (Copi & Burgess-Jackson, 1996).

Deductive arguments are supposed to be inherently logical. This means that, if the premises are true, the conclusion follows necessarily. In other words, if the premises are true, so is the conclusion. Therefore, deductive arguments can either be valid or invalid.

They are either invalid if one of the premises is false (note that the conclusion can still be true in this case, however, the argument would still be invalid due to its deductive nature), or if principles of deductive inferences are violated. One such principle would be that the conclusion necessarily follows from premises in such arguments. Hence, a conclusion and a premise must never be the same.

Inductive arguments are not necessarily conclusive. Instead, premises support conclusions of inductive arguments to some extent. Thus, unlike deductive arguments, inductive arguments are not evaluated as valid (true) or invalid (wrong), but rather with regard to the degree of support for their conclusion which is provided by its premises (Copi & Burgess-Jackson, 1996).

The categorization of fallacies (formal and informal fallacies, see chapter 3.2) is related to the question whether an argument is deductive or inductive.

Third, engaging in argument evaluation can be a complex endeavor. This endeavor involves *four different tasks*: (a) Identification, (b) analysis, (c) evaluation and (d) invention (Walton, 2009).

Identification refers to identifying claims and conclusions (thereby distinguishing them from each other). *Analysis* refers to the task of making implicit claims explicit. *Evaluating* an argument means to judge its strength or weakness based on certain criteria (see chapter 3.2). *Invention* refers to the construction of new arguments.

These four tasks are all relevant skills regarding the recognition of fallacies (i.e., the identification of fallacies as well as explaining why an argument is fallacious).

Taken together, arguments are prone to numerous fallacies and fallacious arguments are a threat to sound decisions that are likely to contribute to solve clients'

problems (Gambrill, 2019; see p. 213 for a list of examples of fallacies). Thus, it is important for social workers to be able to appropriately carry out the four argumentation tasks (i.e., identification, analysis, evaluation and invention) to identify and explain fallacious arguments within social work decision-making contexts. To carry out the tasks argument *evaluation* and *invention*, it is important to be aware of several criteria for good argumentation, against which arguments can be evaluated in order to identify fallacies (Damer, 2011). The concepts of fallacies as well as criteria for good argumentation are explained in the next section.

3.2 Fallacies and Criteria for Good Argumentation

In his theory *Attacking Faulty Reasoning,* Damer (2011) defines a fallacy rather broadly as an error in reasoning or an argument that does not necessarily appear as erroneous. Fallacies can play an important role in decision-making processes with regards to the question whether an argument is accepted or rejected (Tarnoff, 2010). The former argument example (Harold smoking pot, chapter 3.1) might be part of a decision-making process regarding the question what treatment would be best for a clients' problem. Decisions for the implementation of a certain treatment depend to a great extent on the criteria that social workers use to evaluate the accuracy of theories and claims of effectiveness (Gambrill, 2013). Thus, if social workers fail to evaluate claims accurately, they might suggest suboptimal or even harmful treatments to their clients (Gambrill, 2011). The fallacy in the Harold example is called "*post hoc ergo propter hoc*", meaning that whatever precedes another event is the cause of that event. This assumption is implied in the argument and it bridges the gap between the stated claims and the conclusion. As stated before, this assumption is not always true, hence potentially fallacious. If it is fallacious and the fallacy remains undetected, a social worker might

erroneously recommend Harold to break up with his wife. Therefore, and due to the potentially harmful consequences of undetected fallacies, social workers need to be able to identify, analyze and evaluate arguments with regard to their validity or strength of their support for their conclusions and to invent new arguments that explain why a fallacious argument is fallacious. To conduct these tasks appropriately with respect to fallacies, it is helpful to know (1) how fallacies are classified and (2) criteria against which arguments can be evaluated.

The *classification of fallacies* is related to the deductive or inductive nature of arguments (chapter 3.1), in that fallacies are often categorized either as (1) formal or (2) informal (Copi & Burgess-Jackson, 1996).

A *formal* fallacy occurs, when an argument is structurally flawed (i.e., when a conclusion follows not necessarily from one/or a set of claim/s) such as in the following example:

"*If Jim is a dog, then Jim is an animal. (claim)*
Jim is not a dog. (claim)
Therefore, Jim is not an animal." (conclusion)

In this example, the conclusion (Jim is not an animal) does not necessarily follow from the claims, because Jim could of course be any other animal than a dog. Therefore, the argument is fallacious, even though the conclusion could be true.

Informal fallacies refer to arguments that are flawed with regards to their content (i.e., the information provided by their claims) instead of their structure. Consider the following example:

"*I believe in the sanctity of marriage (claim)*

Thus, I think marriage should be between a man and a woman." (conclusion)

This statement by the former US president George W. Bush provides no reason for the connection between the concept of sanctity and the implicit conclusion that only certain gender combinations should be allowed to get married. It is thus flawed with regards to its content.

It can be difficult to reveal such fallacious arguments since they often do not appear to be fallacious. However, in a more fine-grained attempt, Damer (2011) defines fallacies as a violation of one of the following *five criteria for good argumentation*: (1) structure, (2) relevance, (3) acceptance, (4) sufficiency and (5) rebuttal. Evaluating arguments against these five criteria can help to determine whether an argument is fallacious or not. If an argument violates one or more of the criteria, it can be considered fallacious.

An argument violates the *structure* criterion if it is structurally flawed (i.e., the conclusion does not necessarily follow from the claim/s). An example is „The college is very paternalistic in its structures, because it treats students like children." In this example, the claim is the same as the conclusion.

The criterion *relevance* is violated when an argument provides a claim that is not relevant to the conclusion. An example is „Since tobacco can cause cancer, and because it is expensive, specific population groups should not be targeted by tobacco advertisement." The claims might provide good reasons to avoid smoking or not to

advertise for tobacco at all. But they are not specifically targeted at the conclusion that advertisement for tobacco should be restricted to certain populations.

The criterion *acceptance* is violated if a claim fails to meet the standards for acceptability (i.e., linguistic confusion or unwarranted assumptions). An example is „Do you know how to get to the next gas station? Yes, turn right here [driver turns right]. Oh, I did not mean to turn right! Didn't you see that I pointed to the left?" The verbal directions in this example were unclear, since the word *right* – which appears to be a key word with regards to the meaning of the whole argument – can have two very different meanings.

The criterion *sufficiency* is violated if the conclusion is not sufficiently supported by the claims. An example is „Since women in my office do not complain about receiving less money compared to males, they must be satisfied with receiving less pay than their male counterparts." This argument entails the intention to use absence of evidence as actual evidence. Therefore, the conclusion is not sufficiently supported.

The criterion *rebuttal* is violated, if an argument attacks one's opponent in a personal way. An example is „Since you are a man, you cannot speak from a women's perspective about abortion. You have no right to talk about this topic." Here, instead of addressing the issue at hand, the opponent is attacked in a personal way with the intention to lower his credibility.

Knowledge about how to build arguments is especially important for the first two tasks of argument evaluation, i.e., *identifying* claims and conclusions and *analyzing* them to identify implicit claims as necessary prerequisites to appropriately evaluate an argument (chapter 3.1). Knowledge about Damer's (2011) criteria for good argumentation is especially relevant for the third and the fourth task of argument

evaluation. The third task refers to the *evaluation* of the identified conclusion and (implicit) claim/s with respect to their validity (deductive argument/formal fallacies) or to the extent to which a conclusion is supported by means of reasonable or evidentiary claims (inductive arguments/informal fallacies). The skills that refer to the first three tasks of argument evaluation (i.e., identification, analysis and evaluation), are labeled as *fallacy identification*[2] in the context of fallacy recognition. In this context, the fourth task of argument evaluation, *inventing* new arguments (chapter 3.1), can be described as creating explanations for the question why a fallacious argument is fallacious by referring to the criteria for good argumentation. It has already been explained why it is important for social workers to be able to identify fallacies in social work decision-making contexts (chapter 3.1). Likewise, it is also important for social workers to be able to explain why an argument is fallacious for at least two reasons: (1) Ethical and (2) practical ones.

For example, regarding *ethical considerations*, a given fallacious argument might be rejected in a social work decision-making context by another social worker simply by saying "No I don't think so". However, in such a case, the burden of proof remains on the person who rejected the argument and failure to fulfill the burden of proof is a violation of a principle of critical discussions (Damer, 2011; Walton, 1995).

Furthermore, such failure to explain why an argument is fallacious might also have *practical* implications. The respective argument might have proposed a final decision regarding the implementation of an intervention for a client. For example, a social worker might discuss a client's situation in a decision-making context with other professionals in an attempt to make a final mutual decision regarding an intervention for a client. If the

[2] In the second empirical study, the term *fallacy identification* was replaced with *fallacy assessment*, since that term provided a closer and more accurate match to the coding scheme that was used to code the data with regards to the skill *fallacy identification*. However, overall *fallacy identification* is considered to be more coherent. Therefore, it is used in Part I and Part III.

social worker provides no explanation for the rejection of a former fallacious argument, she might not convince the other professionals who might then accept the argument and decide accordingly, even if rejecting the argument and consequently arriving at another decision would have been the better choice for the client. The skills that refer to this fourth task, the provision of explanations for why a fallacious argument is fallacious is labeled *fallacy explanation* in the context of this dissertation.

Subsequently, both, *fallacy identification* and *fallacy explanation* are subsumed under the term *fallacy recognition skills.*

3.3 Summary

This chapter outlined the connection between argumentation and decision-making processes in social work as well as the potential threat that fallacious arguments can represent to sound professional decisions. Chapter 3.1 explained how arguments are constructed out of claims and conclusions. Both can be implicit or explicit. Identifying explicit claims and conclusions and analyzing them (i.e., identifying implicit claims and conclusions) are important first steps which have to be undertaken in order to evaluate arguments with regards to their quality (i.e., either the validity of their structure or the strength of the support for a conclusion provided by one or more claims). Chapter 3.2 outlined the concept of fallacies. It was argued that a fallacy is a violation of one of five criteria for good argumentation. Each of these criteria entails different principles. It is important for social workers to know these principles and to be able to apply them in order to appropriately evaluate arguments and to explain why fallacious arguments are fallacious. In the context of this dissertation and with respect to fallacy recognition, the identification, analysis and evaluation of conclusions and claims is called *fallacy identification*. The ability to invent new arguments with the goal to explain why a

fallacious argument is fallacious is called *fallacy explanation*. Both skills are summarized with the *term fallacy recognition skills*.

4. Fostering Evidence-based Practice and Argument Evaluation in Social Work

Chapter four concerns instructional approaches that are potentially effective in facilitating EBP and/or the use of ESTs (which is especially relevant for the first empirical study; chapter 6) as well as a specific instructional approach and scaffolds (i.e., a kind of support that helps the learner to learn through activity; Tabak, 2004) that are promising means to facilitate fallacy recognition skills (which are especially relevant for the second empirical study; chapter 7).

Chapter 4.1 addresses two groups of instructional approaches that differ in the extent to which they provide learners with freedom regarding the question how to structure their learning process. Approaches that grant learners more freedom are called *student-centered* (chapter 4.1.1) and those with less degrees of freedom are called *teacher-centered* (chapter 4.1.2). This distinction is particularly relevant for the first empirical study (chapter 6). Chapter 4.2 outlines a student-centered instructional approach, namely *Case-based Reasoning*, that is of special importance for the second empirical study (chapter 7). Chapter 4.3 introduces two scaffolds called *worked examples* (chapter 4.3.1) and *scripts* (chapter 4.3.2). Such scaffolds provide learners with further support that can enhance their learning outcomes beyond the use of an instructional approach such as Case-based Reasoning.

4.1 Teacher- vs. Student-Centered Approaches

In general, an important question for the design of a learning environment refers to the degree of freedom learners are granted to structure their learning processes (Kollar & Fischer, 2019). *Teacher-centered* (or direct) approaches provide learners with less

freedom, while learners in *student-centered* (also often referred to as constructivist) approaches can organize their learning processes to a greater extent on their own. The respective amount of freedom can vary vastly in real world practice. However, for analytical as well as for didactical purposes, a rather dichotomous distinction is made in the following section with reference to student-centered (chapter 4.1.1) and teacher-centered approaches (chapter 4.1.2; Kollar & Fischer, 2019).

4.1.1 Fostering Evidence-based Practice and Argument Evaluation through Teacher-centered Approaches

Teacher-centered approaches structure learning processes to a higher extent compared to student-centered approaches. They include methods that offer rather direct instruction such as presentations. The basic idea is to have a teacher who explains concepts and procedures to the learners and induces specific practice units (Kirschner, Sweller, & Clark, 2006). An important argument regarding teacher-centered approaches refers to the *human cognitive architecture*, i.e., the structural basis on which humans process information (Sweller, Ayres, & Kalyuga, 2011).

Different models of human cognitive architecture exist. Such models are theoretical attempts that try to explain how the human memory processes information. Theories of the human cognitive architecture can be roughly distinguished between those that assume that human memory consists of only one unit that stores information (e.g., Anderson, 1996) and those that ascribe multiple units for information storage to the human memory (e.g., R. C. Atkinson & Shiffrin, 1968; see Zoelch, Berner, & Thomas, 2019 for a brief overview). Models that assume only one unit for information storage are not

relevant for the following argument. Therefore, only a model that refers to multiple units is explained now.

R. C. Atkinson and Shiffrin (1968) proposed a model with three storage units (i.e., structural features of the memory system): the (1) sensory register, (2) short-term store and (3) long-term store (the more casual terms short-term and long-term memory were avoided to express the idea that these different units resemble a store with different capacities; Hoffmann & Engelkamp, 2017).

When a human perceives a stimulus from the external environment, the very stimulus enters at first the *sensory register*. Therefore, the sensory register has to process huge amounts of information. Consequentially, it must have a great capacity. However, it is assumed that it holds information only for a very short time (presumably 0.5 to 1 second(s); Sperling, 1960). The function of the sensory register is to filter relevant information and transfer it to short-term store.

The *short-term store* is considered to be an interface that processes information in a serial manner between the sensory register and the long-term store. It is assumed that the short-term store has a very limited capacity (e.g., Cowan, 2001, see also below) and that it can hold information only for a few seconds. However, this duration can be increased through rehearsal. The longer novel information remains in short-term store, the more likely this information will be transferred to long-term store.

Long-term store is supposed to have a potentially infinite capacity to store information. It is also assumed that it can hold information for an infinite amount of time (R. C. Atkinson & Shiffrin, 1968).

The theory of R. C. Atkinson and Shiffrin (1968) can be considered an important theoretical model for empirical research, however, it was especially criticized for the idea

that information is processed in a serial manner as well as for the subsequent proposal that rehearsal makes it more likely that information becomes stored in long-term store (Hoffmann & Engelkamp, 2017). This idea is not convincing, since some information is stored in long-term store, even when it is not subject to rehearsal, such as experiences that cause strong emotional arousal (McGaugh, 2013). Subsequently, Baddeley and Hitch (1968) proposed a different model for the short-term store, known as the *working memory*. The concept of *chunks* can help to understand an important difference between the idea of a short-term store and a working memory. Miller (1956) distinguished *bits* of information and *chunks* of information with chunks being essentially a group of bits. While the capacity of short-term store and/or working memory is very limited to a certain number of chunks (e.g., four plus/minus one according to Cowan, 2001 or seven plus/minus two according to Miller, 1956), chunks themselves can hold different amounts of bits (their capacity may for example increase with increasing expertise; Chase & Simon, 1973). An example for bits is the following sequence of letters with each letter being one bit: "l – o – l – r – o – f – l – e – g – i – e". This sequence could be transformed within the short-term store (or more precisely: the working memory) into the following four chunks: "lol – rofl – e.g. – i.e.". This process however ascribes an *active* role to the short-term store. It must integrate novel information (the sequence of letters) in already available semantic information from long-term store (i.e., prior knowledge). It is thus a *working* memory.

Proponents of rather teacher-centered instruction refer to a model of human cognitive architecture that is based on working memory and long-term memory (in this doctoral thesis, this term is used synonymously with the already introduced long-term store). The long-term memory holds *schemas* (i.e., cognitive structures) that combine

multiple units of information into a single unit in such a way, that using the single unit resembles the usage of all multiple units. For example, the equation $a/b = c$ can be solved by multiplying both sides with the denominator *b.* A person who acquired this schema is likely to resolve this equation and all other problems with a similar structure in only one step. This allows for efficient problem-solving processes, since schemas can be applied and executed unconsciously after a sufficient amount of practice. Therefore, altering long-term memory in the sense of schema acquisition is the goal of all learning processes (in other words, if the long-term memory has not been altered by schema acquisition, nothing has been learned; Kirschner et al., 2006). The interface between external information and the long-term memory is the working memory. It works in both directions: On the one hand, it processes and transmits novel external information. On the other hand, it processes information from the long-term memory to ensure that this information is used appropriately in a given environment. However, while working memory can process a lot of information that is stored as schemas in the long-term memory at once, it is very limited in the amount of novel information that has to be processed. This is because novel, random and unstructured information must be processed consciously, which takes cognitive effort. The argument that follows from such a model of human cognitive architecture is that instruction with the goal to facilitate learning (i.e., long-term memory alteration) should present the novel information that should be learned in a way that ensures that the working memory is not overstrained. Proponents of this argument assert that student-centered approaches overstrain working memory in problem-solving contexts, because they require the learner to process information that is both, relevant and irrelevant for learning processes (as compared to teacher-centered approaches that might provide only relevant information). For instance, student-centered approaches might trigger learning strategies such as

means-ends-analysis which requires a learner to search the problem space for information that would otherwise have been provided directly by the teacher herself. As such, strategies that put a huge demand on the working memory make schema acquisition (i.e., learning) more unlikely (Kirschner et al., 2006; Sweller et al., 2011). Hence, it is assumed that rather teacher-centered approaches provide a better fit to the human cognitive architecture, for example by pointing the learner's attention to information that is relevant for the learning process. That way, processing novel information is supposed to stay within the limits of working-memory, which should make schema acquisition more likely (see also chapter 4.3.1.2 on Cognitive Load Theory for a more elaborated theoretical account on schema acquisition).

An example for such a teacher-centered approach is *direct instruction* (Slavin, 2018). Here, Slavin (2018) suggests 7 steps: (1) Formulate learning goals. This helps learners to focus on aspects that are relevant for learning. (2) Activate learners' prior knowledge, for example by asking questions. This should make it more likely that learners integrate new information with pre-existing knowledge structures relevant to the learning process. (3) Present novel information in a structured, clear and efficient way. This should prevent the working memory from becoming overstrained. (4) Use comprehension checks (such as questions) while novel information is provided. This can reveal knowledge gaps which have yet to be closed. (5) Let learners apply the knowledge from previously presented information. This should help learners to transfer knowledge to new situations. However, in early learning stages, learners should receive further support when they work on their own (e.g. with worked examples, see chapter 4.3.1) before they work independently (van Merrienboer, 2013). (6) Encompass further elaboration such as homework. This can further facilitate schema acquisition. (7) Assess

performance and give feedback. This helps learners to keep track of their own learning processes, which may in turn trigger new ones. Renkl (2009) suggests that such instructional approaches have the highest prevalence in German schools. However, it is important to deliver direct instruction appropriately which is often not the case (Kollar & Fischer, 2019).

Hattie (2009) reports in his meta-meta-analysis an average effect size of $d = .59$ (moderate) for direct instruction compared to other traditional instructional approaches. There is ample empirical evidence from experimental studies that suggests the superiority of teacher-centered approaches such as direct instruction compared to rather student-centered ones such as mere problem-solving (e.g., Kyun, Kalyuga, & Sweller, 2013; Oksa, Kalyuga, & Chandler, 2010; Paas & van Merrienboer, 1994; Rourke & Sweller, 2009; Sweller & Cooper, 1985). However, this superiority decreases with increasing prior knowledge (e.g., Blayney, Kalyuga, & Sweller, 2015; Khacharem, Zoudji, & Ripoll, 2013; H. Lee, Plass, & Homer, 2006; Lorains, Ball, & MacMahon, 2013; Reisslein, Atkinson, Seeling, & Reisslein, 2006). This is known as the *expertise-reversal-effect* (Kalyuga, 2007; Kalyuga, Rikers, & Paas, 2012; Khacharem, Zoudji, & Kalyuga, 2015; see also chapter 4.3.1.2). Furthermore, knowledge acquired through teacher-centered approaches often remains inert (Renkl, Mandl, & Gruber, 1996). Such knowledge is hardly accessible for problem-solving purposes. One explanation for this problem is that knowledge might be context-bound and that it is therefore difficult for learners to transfer knowledge acquired through teacher-centered learning approaches to new situations (Barnett & Ceci, 2002). Some argue that student-centered approaches might be better suited to address this problem of transfer of acquired knowledge to new situations (cf., Renkl et al., 1996). Such approaches are explained in the next chapter.

4.1.2 Fostering Evidence-based Practice and Argument Evaluation through Student-centered Approaches

In *student-centered approaches*, learners structure their learning processes to a greater extent on their own, compared to teacher-centered approaches (Kollar & Fischer, 2019; Renkl, 2009). Student-centered approaches usually present learners authentic, meaningful and more complex problems, which they are supposed to solve either alone or as a group. Providing meaningful problems together with a high degree of freedom regarding the structure of one's own learning process is supposed to lead to a high level of learners' engagement in the learning processes, thereby facilitating not only the acquisition of domain-specific knowledge, but also the acquisition of heuristics (i.e., cognitive strategies that ignore part of the information; Gigerenzer & Gaissmaier, 2011), meta-cognitive knowledge (i.e., knowledge regarding the relation of declarative and procedural knowledge[3], such as – for example – learner's self-monitoring of learning processes; Veenman, van Hout-Wolters, & Afflerbach, 2006) and learning strategies (i.e., behavioral and cognitive approaches to acquiring knowledge, such as mind mapping; Mandl & Friedrich, 2006).

An example for a student-centered instructional approach is *Problem-based Learning* (PBL; Barrows & Tamblyn, 1980; Hmelo-Silver, 2004). PBL was initially developed in the medical field (Barrows & Tamblyn, 1980). In PBL, students solve authentic problems in groups. *Situated learning* (Lave & Wenger, 1991) contributed to the further development of PBL (Kollar & Fischer, 2019). According to situated learning it is easier for learners to transfer the application of knowledge to new situations when they acquire the respective knowledge in authentic contexts. PBL can be seen as an

[3] Declarative knowledge refers to knowledge about concepts and principles, whereas procedural knowledge refers to knowledge about the application of declarative knowledge (e.g., Anderson, 1996).

umbrella term for various methods that share the same core aspects (e.g., authentic problem, discuss the problem in groups to activate prior knowledge before receiving further instruction, discover knowledge gaps to trigger interest), but differ with regards to certain aspects such as learning material. For example, in PBL as proposed by Hmelo-Silver (2004), student groups put their ideas and hypotheses regarding the solution of the problem at hand on a specially formatted whiteboard. Afterwards they divide up learning issues, examine them individually and then get back together to reflect on the new knowledge in order to come up with new hypotheses and to proceed in the problem-solving process. This learning cycle continues until a satisfying level of satiation is accomplished.

Hattie's meta-meta-analysis reports a rather small effect of PBL on learners' learning achievements compared to instruction that is not problem-based ($d = .15$). However, it suggests various moderators such as prior knowledge. Several other meta-analyses suggest the effectiveness of student-centered approaches such as PBL (Schmidt et al., 2009, Dochy, Segers, van den Bossche, & Gijbels, 2003 Gijbels, Dochy, van den Bossche, & Segers, 2005) or Inquiry-based Learning (Lazonder & Harmsen, 2016), which is largely comparable to PBL (Hmelo-Silver, Duncan, & Chinn, 2007). The meta-analysis by Schmidt et al. (2009) suggests that PBL is particularly more effective compared to more traditional approaches regarding skills and application-oriented knowledge and that its effectiveness ascends with increasing expertise of the learners. This is in accordance with the expertise-reversal effect that may occur in rather teacher-centered approaches with learners' increased prior knowledge.

There has been considerable debate about the question which approach is generally preferable for instructional purposes (Hmelo-Silver et al., 2007; Kirschner et

al., 2006; Kuhn, 2007; Schmidt, Loyens, van Gog, & Paas, 2007; Sweller, Kirschner, & Clark, 2007). As stated before, the main argument of proponents of teacher-centered approaches is that student-centered approaches do not fit the human cognitive architecture (i.e., these approaches confront the learner with too many tasks and information, which is supposed to overstrain the capacity of their working memory; Kirschner et al., 2006), especially if learners have little prior knowledge that could help them organize the tasks and information. Hence, proponents of teacher-centered approaches argue that student-centered approaches are less effective (Kirschner et al., 2006), whereas advocates of student-centered approaches emphasize that various aspects such as the goal of the learning process (e.g., the acquisition of declarative vs. procedural knowledge; Renkl, 2009), the context (i.e., what is being taught; Kuhn, 2007), or learners' prior knowledge (cf., Kollar & Fischer, 2019) have to be taken into account when it comes to the question to which extent learning environments have to be structured. Furthermore, they argue that Kirschner et al. (2006) misinterpreted student-centered approaches in the sense that such approaches – even though they stress the importance of self-directed learning processes – often offer a considerable amount of *guidance* (Hmelo-Silver et al., 2007).[4] In fact, empirical evidence suggests that guided student-centered approaches are superior to (student-centered) approaches that engage learners in pure discovery (Mayer, 2004). For example, the meta-analysis by Lazonder and Harmsen (2016) provides strong evidence suggesting that Inquiry-based Learning is especially promising when learners are provided with additional guidance. Such guidance can be provided through *scaffolds*. Scaffolds are means that provide support

[4] Furthermore, there have been recent attempts by authors such as Loibl, Roll, and Rummel (2017) or Sweller and Paas (2017) to overcome the rather dichotomous view of instructional approaches by thinking about ways to integrate both paradigms.

that helps the learner to learn through activity (Tabak, 2004; see also chapters 4.3.1 and 4.3.2).

Taken together, empirical evidence suggests that both approaches, teacher- as well as student-centered ones, can be beneficial for learners' learning processes. It is important to consider various aspects such as the learning goals or learners' prior knowledge as well as the phase of knowledge- and/or skill-acquisition. Teacher-centered approaches seem to be more promising at early phases, while student-centered approaches are especially effective in later stages and for skill-acquisition. Moreover, they can benefit from additional guidance such as scaffolds.

The process of Evidence-based Practice, the use of Empirically Supported Treatments as well as fallacy recognition skills can all be considered *skills* that have to be applied in different (and thus transferred to) meaningful contexts *in practice* which is – as outlined before – considered to be a particular strength of rather student-centered approaches. Therefore, it is reasonable to assume that a rather student-centered approach is preferable in facilitating learners' acquisition of these skills. *Case-based Reasoning* is an example of a student-centered approach that can be considered a promising means to foster social work students' fallacy recognition skills. It is described in the following chapter.

4.2 Case-based Reasoning (As an Example for a Student-centered Instructional Approach)

The development of PBL in medicine was closely related to the fact that medical students experienced problems when they were supposed to apply knowledge derived from books in diagnostic situations. This circumstance also led to increasing requests for

instructional approaches that induce knowledge application based on authentic cases, rather than simply learning from books (Gräsel & Mandl, 1993). Case-based Reasoning (CBR) is a related instructional approach (e.g., Kolodner, Camp et al., 2003).

CBR is rooted in work by Roger Schank. Schank's (1999) model of dynamic memory suggests that humans understand events based on their knowledge structures (see also 4.3) and prior experiences. According to Schank, experiences provide us with the knowledge (i.e., knowledge structures) that is most important to live our daily lives. The basic premise is that when we encounter a new and unfamiliar event, we have to rely on previous experiences that help us understand and act in the situation. By reminding us of these helpful experiences and applying them to the unfamiliar situation, related knowledge structures are updated accordingly (hence the term *dynamic* memory). Thus, "reminding is the process by which Case-based Reasoning (e.g., reasoning by relying upon cases of experiences we have previously stored in our memory, rather than reasoning by applying rules) takes place" (Schank, 1999, p. 174). This implies the notion that problem-solving activities from a CBR-perspective rely on knowledge that is not necessarily rationally known by the problem solver. Subsequently, CBR was put forward as an instructional approach in the 1990's, especially by the group around Janet Kolodner (Kolodner, 1993).

In CBR, learning means to interpret new experiences, integrate these interpretations into the memory and reindex old experiences in order to make them more accessible for future problem-solving. Interpreting new experiences means to find explanations that connects goals, actions and outcomes. Therefore, CBR entails two important aspects: (1) integrating experiences (cases) into the memory (i.e., assigning indexes or labels to these memories to "find" them as quickly as possible) and (2) recalling the most appropriate case for a given problem to solve it by the time one is

confronted with it. These two aspects of memory integration of cases and appropriate recall in a problem-solving situation refer to the so-called *indexing problem* (Kolodner, 1992). It is important for the problem solver to evaluate the problem in enough detail in order to identify the index of the most appropriate case. The more willing the problem solver is to explore the problem, the more likely she is to find the most appropriate case (Kolodner, Camp et al., 2003). Thus, "motivation, opportunity, and ability to explain are key to promoting learning" (Kolodner, Camp et al., 2003, p. 502).

Accordingly, CBR entails four important claims for effective instruction: (1) authentic cases provide a better (i.e., more memorable and motivating) opportunity for the learner to provide explanations, (2) unsuccessful reasoning shows the learner what needs to be learned, which implies the necessity to provide the learner with opportunities to fail, (3) reflection on cases may enable the learner to avoid failures in the future and (4) learners may receive educational assistance to identify those explanations that work best for future problem-solving (Kolodner, Cox, & Gonzalez-Calero, 2005).

Empirical evidence regarding the effectiveness of CBR of studies in social work is rare, but there is evidence coming from fields other than social work. For example, quasi-experimental studies with middle school students ($N = 240$) suggest that case-based approaches, tools and activities are more effective in facilitating students' scientific thinking and peer collaboration skills than rather teacher-based learning environments (Kolodner, Camp et al., 2003; Kolodner, Gray, & Fasse, 2003). In another quasi-experimental study conducted in South Africa, Postma and White (2016) found CBR to be more effective in the facilitation of clinical reasoning skills (e.g., diagnostic competencies) among dental students ($N = 136$) compared to lecture-based instruction. Hong and Yu

(2017) conducted a randomized controlled trial with undergraduate nursing students ($N = 122$) and found different styles of case-based lectures to be effective in the facilitation of critical thinking. Harman et al. (2015) investigated qualitative data from dietetics students ($N = 85$). The students perceived competency gains in critical thinking and problem-solving skills through CBR. A literature review by Popil (2011) concludes that CBR supports the development of critical thinking skills among nurses and other health care professionals.

Taken together, empirical evidence suggests that CBR is a promising means to facilitate skills such as scientific thinking, clinical reasoning, critical thinking and problem-solving. All these skills are closely related to Evidence-based Practice and fallacy recognition skills, as outlined in chapters 2 and 3 (see also Gambrill, 2013). Therefore, it is reasonable to assume that CBR is a promising means to facilitate these skills. The effectiveness of a student-centered approach such as CBR is likely to be enhanced through the additional use of scaffolds as already mentioned in chapter 4.1.2. *Worked examples* as well as *scripts* are scaffolds that can be considered promising means to enhance students' learning outcomes beyond the effects of CBR. Both scaffolds are introduced now.

4.3 Worked Examples and Scripts as Additional Scaffolds for Case-based Reasoning

Scaffolds were initially described as some sort of tutor (e.g., an adult who models the parts of a problem-solution process that are beyond the capacity of a child; Wood, Bruner, & Ross, 1976). Nowadays a scaffold is considered a kind of support that helps the learner to learn through activity (Tabak, 2004). However, the basic idea of a scaffold has not

changed: To support learners during active (i.e., learners apply declarative knowledge that they have learned upfront in order to proceduralize it) learning phases in such a way that they can successfully solve problems they would not be able to solve without the respective support (Quintana et al., 2004). Worked examples and scripts are two such scaffolds that can be applied during CBR. This chapter introduces worked examples (chapter 4.3.1) and scripts (chapter 4.3.2).

4.3.1 Worked Examples to direct Learners' Attention to Information that is relevant for Schema Acquisition

Worked examples are scaffolds that can be successfully integrated in CBR (e.g., Kopp, Stark, & Fischer, 2008). Chapter 4.3.1.1 provides a definition of worked examples as well as a brief overview of their effects. There is a vast amount of empirical evidence suggesting the effectiveness of worked examples in comparison with mere problem-solving. The respective worked-example effect is typically explained with Cognitive Load Theory which is outlined in chapter 4.3.1.2. Initially, worked examples were especially used in well-structured domains such as mathematics. However, they are increasingly applied in ill-structured domains such as medicine. It is important to consider their application in such domains in the context of this thesis, since social work can be considered a rather ill-structured domain. Their use in such domains as well as respective empirical evidence is illustrated in chapter 4.3.1.3.

4.3.1.1 Definition and Potential Effects

A worked example is an instructional method which is especially aiming for cognitive skill acquisition (Renkl, 2011). Typically, it entails three components: (1) the formulation of a problem, (2), steps that have to be undertaken to arrive at the correct solution (however,

these are sometimes missing), and (3) the correct solution (Renkl, 2014). Usually, learners are provided with information on concepts and principles (e.g., the concept of subtraction in math) and afterwards, they work with worked examples (e.g., a concrete equation that resembles all the steps that have to be undertaken to achieve the correct solution) in order to transform the information into declarative and procedural knowledge.

A vast amount of empirical evidence shows that such worked examples are more effective in facilitating cognitive skills than mere attempts of problem-solving (e.g., Cooper & Sweller, 1987; Eysink et al., 2009; Hilbert, Renkl, Kessler, & Reiss, 2008; Sweller & Cooper, 1985; see R. K. Atkinson, Derry, Renkl, & Wortham, 2000 and Renkl, 2005 for overviews). For example, in their seminal study, Sweller and Cooper (1985) ran a series of experiments to investigate the effects of worked examples on high school students' algebraic problem-solving skills (dependent variables were completion time and number of mathematical errors). For instance, in experiment 3, 22 ninth grade high school students were provided with worked examples of algebraic problems. Then, students in the control condition (conventional problem-solving) worked on eight algebraic problems that represented four different problem types. The students in the experimental condition (with worked examples) worked on the same problems. However, the first problem of each problem type was worked out. Students of the experimental condition needed less time and committed fewer errors in a post-test with four structurally identical problems than students of the control condition.

While it is true that this so-called *worked-example effect* was initially mostly observed in studies that offered little to no support for the control group (Sweller et al., 2011), there is growing empirical evidence suggesting that worked examples are also

superior to problem-solving when problem-solving is supported (Salden, Aleven, Schwonke, & Renkl, 2010; Schwonke et al., 2009). For example, Schwonke et al. (2009) conducted a study to investigate the effects of worked examples on students' procedural skills (i.e., ability to solve numerical problems in circles geometry) and conceptual understanding (i.e., ability to apply acquired principles not in calculations but in explanations, arguments and evaluation). In experiment 2, 22 eighth grade students and 28 ninth grade students from a German high school were randomly assigned to either an experimental (with worked examples and with a cognitive tutor) or a control (with a cognitive tutor) condition. Students worked on seven problems regarding geometric principles. The tutor provided just-in-time feedback, hints and production rules. The worked examples provided solution steps as well as a correct solution. In a post-test, students of the experimental condition needed less learning time and showed a deeper conceptual understanding of principles relevant to solve geometry problems than students of the control condition.

However, worked examples can also yield negative effects, for example due to suboptimal design. Examples for such negative effects are the (1) split- attention effect and the (2) redundancy effect.

A potentially negative effect that should be considered when designing worked examples is the so-called *split-attention effect*. Split-attention (i.e., a condition) "occurs when learners are required to split their attention between at least two sources of information" (Sweller et al., 2011, p. 111). For example, split-attention occurs when a student has to read a set of equations that refer to a geometrical diagram (Tarmizi & Sweller, 1988). Split attention can yield a negative effect, due to the need for the learner to switch attention between different sources of information which puts demand on the

working memory, thereby leading to increased cognitive load (i.e., it stresses the limited capacity of the working memory; see also chapter 4.3.1.2).

An important question regarding the avoidance of the split-attention effect is whether each of the different learning materials is relevant for the learning process but at the same time intelligible in isolation (i.e., it cannot be fully comprehended when it is presented exclusively). In such a situation, it would be necessary to present the learning materials in an integrated way to relief the learner from the necessity to split her attention (Sweller et al., 2011).

Much empirical evidence suggests that split attention conditions put heavy demands on the working memory as well as the existence of the split-attention effect (e.g., Bauhoff, Huff, & Schwan, 2012; Mayer, Steinhoff, Bower, & Mars, 1995; Owens & Sweller, 2008; Pociask & Morrison, 2008; Tarmizi & Sweller, 1988). For example, Pociask and Morrison (2008) conducted a study to investigate the effectiveness of instructional materials to control for split attention regarding the facilitation of orthopedic physical therapy skills. 41 physical therapy students were randomly assigned to either an experimental (modified instruction) or control condition. The participants of both conditions received a lecture on back pain localization and evaluation. However, the control group instruction included split-attention features (e.g., a photograph of a physician performing a task with referring body text). In the experimental group, respective material was presented in an integrated way. In post-tests, students of the experimental group scored significantly higher on a knowledge test as well as on psychomotor tasks and reported lower levels of cognitive load on both tasks.

Another example for a negative effect that can be yielded by worked examples is the *redundancy effect*. Just like the split-attention effect, it can result from a split attention

condition. The difference is that in the case of the redundancy effect, each of the different informational sources entails the same information that could also be understood independently. Consequentially, one informational source would already be enough to deliver all the information that is necessary for a problem-solving process. In such a case, additional information that is redundant for schema acquisition has yet to be cognitively processed by the learner. In contrast to the split-attention effect, the information of these different sources cannot be integrated. Instead, only one source that provides all the information that is required to understand and solve a problem should be presented and no further information should be added to avoid the redundancy-effect (Sweller et al., 2011).

Numerous studies found empirical evidence for the redundancy effect (e.g., Cerpa, Chandler, & Sweller, 1996; Chandler & Sweller, 1996; Pociask & Morrison, 2008; Sweller & Chandler, 1994). For instance, Cerpa et al. (1996) conducted an experiment (exp. 1) in which they randomly assigned a group of ninth grade students of an all-male Australian high school ($n = 32$) to either (1) a group using a computer-based training or to (2) a group using the same computer-based training plus a hard-copy of the training materials. The students went through an instructional phase in which they used the instructional materials to learn how to use the software. In a subsequent test-phase, students of the computer-based training group outperformed students of the computer-based training group with the additional hard copy on performance tasks.

Effects of worked examples can be explained with the Cognitive Load Theory (Sweller et al., 2011), which is the classic approach to explain the effectiveness of worked examples (Renkl, 2014). It is described in the next section.

4.3.1.2 Cognitive Load Theory

The Cognitive Load Theory (CLT) is an attempt to explain how learning processes work with the goal to derive instructional principles from these explanations (Sweller et al., 2011). The basis for the explanations refers to a specific model of the human cognitive architecture (see chapter 4.1). A basic assumption of CLT is that the human cognitive architecture is predominantly defined by working memory and long-term memory (Kirschner et al., 2006). *Long-term memory* holds the knowledge base for any cognitively based activity. Thus, learning in this sense means to alter long-term memory, that is, to transfer novel information from working memory to long-term memory (known as schema acquisition). *Working memory* is the structure for conscious processing. In contrast to long-term memory, we are consciously aware of any information that is processed in working memory. But, working memory has a very limited capacity. Miller (1956) proposed a capacity of seven +/- two chunks (see also chapter 4.1) and more recent work declares even stricter assumptions. For example, Cowan (2001) proposes a capacity of only four +/- one chunks. Moreover, information that is not subject to practice is lost quickly. This is however not the case for information that is stored in long-term memory and brought back to working memory for practical reasons. Hence, from this perspective, any kind of instruction must consider the limitations of the working memory (Kirschner et al., 2006). These limitations are crucial in situations where a learner encounters a problem for which she holds no relevant knowledge (i.e., schema) in long-term memory. Because then, she must use working memory to search the problem space for problem-relevant information. Since this search relates solely to novel information, it puts heavy demands on the working memory, thereby making schema acquisition more unlikely. Demands on working memory are called *cognitive load*. In CLT, three different kinds of cognitive load are distinguished: (1) Intrinsic load, (2) extraneous load and (3)

germane load (see de Jong, 2010 for a critical discussion regarding the question whether these three kinds of cognitive load can be clearly distinguished).

Intrinsic load is caused by cognitive processes that are necessary to understand and process the task at hand. Sweller et al. (2011) conceptualize it as the number of elements a learner must deal with in order to complete a task. This number depends on the learners' prior knowledge as well as on the complexity of the task at hand. For example, a first-year social work student is likely to have no prior knowledge of theories regarding social diagnosis. Such theories emphasize the importance of clear and objective descriptions of relevant information regarding a social problem, thereby also illustrating the nature of terms and phrases that lag such clear descriptions. An example for such a term might be a *typical war veteran*. If a first year social work student encounters the task to evaluate the argument 'the client is a typical war veteran, therefore, we should put him in a self-help group', the student might start to search the problem space for potential attributes of a typical war veteran in order to evaluate these attributes against necessary prerequisites for the mediation of a client to a self-help group. A more advanced student who is experienced with social diagnosis theories might immediately recognize the problematic phrase and reject the argument due to vague language. Therefore, the same task poses different amounts of intrinsic load to different learners due to their prior knowledge. Thus, intrinsic load cannot be changed by instruction (however, efforts such as sequencing to-be-learned information from simple to complex, so that a learner does not perceive its full complexity at once, might be considered as an attempt to reduce intrinsic load; de Jong, 2010; van Merrienboer, Kirschner, & Kester, 2003).

Extraneous load is caused by an engagement in cognitive processes that are not directly relevant for schema acquisition. The more elements that are unnecessary for completing a task are provided by a learning environment, the more cognitive resources

must be used to take them into account simultaneously, which makes schema acquisition more unlikely. Regarding the example with the typical war veteran, a potential learning environment that is supposed to help a student to identify the argument as fallacious and to reject it accordingly could, for example, provide the student with an elaborated text on social diagnosis theory. Or it could simply provide an example of a similar argument that also uses vague speech and that is known to be fallacious due to the use of such language, thereby pointing the learners' attention to the problematic aspect of the argument with the typical war veteran. If such an example provides the learner with all the information that is relevant to solve the task at hand, it is more likely to foster schema acquisition. This is because it entails much less information that has to be processed in working memory (i.e., it poses less extraneous load on the working memory), than a long text on social diagnosis theory.

Germane load can be described as *desirable* load. It is caused by cognitive processes that are related to schema acquisition, such as interpretation, classification or organization of information (Mayer, 2002). Therefore, a basic idea is to keep extraneous load as low as possible to have more free capacity in the working memory for germane load. However, this free capacity is not automatically transformed to germane load. Inducing germane load might be achieved by developing instructional designs that direct the learners' attention to what is relevant for learning, while at the same time withdraw attention from what is not relevant. Furthermore, it is important to ensure that the balance of the three different kinds of load stays within working memory limits. The Journal *Learning and Instruction* dedicated a whole issue (vol. 16, issue 2) to the question how to manage cognitive load to foster transfer. Yet, little is known about the nature of germane load and de Jong (2010) proposed to focus more on the question how to induce germane load in research on cognitive load. It can, for example, be important for schema

acquisition that learners deeply process worked examples by explaining their underlying principles to themselves. So-called *self-explanations* relate to the extent to which a learner explains the rationale (i.e., underlying concepts and principles) of a presented worked example to himself. However, research shows that learners differ in the quality of their self-explanation (Chi, Bassok, Lewis, Reimann, & Glaser, 1989) and that they do not necessarily engage in fruitful self-explanation processes during their work on worked examples (Renkl, 1997). Thus, prompting self-explanations might be one way to induce germane load when working with worked examples (Renkl, 2014, 2017).

Likewise, Paas and van Gog (2006) suggest that strategies that lead to a higher understanding of the solution process may increase germane load when they are combined with worked examples. Subsequently, this could mean that additional support that prompts the learner to *apply* the underlying principles of the worked examples might increase germane load. An example for an activating scaffold is a script (see chapter 4.3.2). In the context of the typical war veteran example, a script could, for instance, prompt the learner to evaluate the argument against the criteria for good argumentation that is related to vague speech (acceptability; Damer, 2011), thereby asking the learner to use the underlying principles of the worked examples.

Taken together, worked examples are commonly attributed to the reduction of extraneous load, because they make unnecessary search processes obsolete. At the same time, working memory capacity can be used to study aspects that are relevant for schema acquisition. In other words, reduced extraneous load leaves cognitive capacities for learning-relevant (germane) load (Paas & van Gog, 2006). However, research suggests that this works better with novices than with experts for whom even negative effects are possible (*expertise-reversal effect;* Kalyuga, 2007; Kalyuga et al., 2012). A possible

explanation for this effect refers to increased extraneous load in a leaner's working memory, which can occur when worked examples point learners' attention to aspects that they already know and are thus redundant for learning. Hence, worked examples are presumed to be especially effective in earlier phases of skill acquisition (Renkl, 2011).

Initially, worked examples were considered to be particularly suited for well-structured domains such as mathematics (e.g., Renkl, 2005). However, social work can be considered to be a rather ill-structured domain, since the solution of social problems requires the application of knowledge from various domains such as social work, sociology, law or medicine (Saub-Bernasconi, 2018). Therefore, it is important in the context of this doctoral thesis to discuss the application of worked examples in rather complex and ill-structured domains. The next chapter will provide further insights on the application of worked examples in such domains.

4.3.1.3 Worked Examples in Ill-Structured Domains

It seems reasonable to assume that worked examples are particularly suited for highly structured domains like mathematics and not for rather complex domains such as social work, since highly structured domains imply straight-forward solution processes which can be modeled with a worked example more easily. For example, Renkl (2005) argued, that the effectiveness of worked examples is more or less restricted to algorithmic skill (i.e., well-structured) domains. Indeed, initially, research on worked examples focused especially on well-structured domains, where example-based learning was frequently found to be superior to mere problem-solving (e.g., Carroll, 1994; Cooper & Sweller, 1987; Paas & van Merrienboer, 1994; Sweller & Cooper, 1985). However, more recent research also applied worked examples successfully in rather ill-structured domains such

as medicine (Kopp et al., 2008; Kopp, Stark, Kühne-Eversmann, & Fischer, 2009; Stark, Kopp, & Fischer, 2011), design (Rourke & Sweller, 2009), literacy (Oksa et al., 2010), English literature (Kyun et al., 2013) and related skills (related in the sense that such skills do not necessarily follow nomothetic rules) such as argumentation (Kollar et al., 2014; Schworm & Renkl, 2007), reasoning on legal cases (Nievelstein, van Gog, van Dijck, & Boshuizen, 2013) or collaboration (Rummel, Spada, & Hauser, 2009). For example, Kollar et al. (2014) conducted an experimental study with pre-service teachers ($N = 101$) with the factors collaboration scripts (with vs. without) and heuristic worked examples (i.e., worked examples that provide not only a solution but also underlying principles; Renkl, Hilbert, & Schworm, 2009) vs. problem-solving to foster mathematical argumentation skills). After an instruction phase in which the pre-service teachers were provided with information on elementary number theory, a pre-test regarding the teachers' mathematical argumentation skills was conducted. During the experiment, pre-service teachers in the problem-solving condition received a mathematical problem from elementary number theory (e.g., choosing an odd amount of consecutive numbers, sum them up, think about unusual aspects, find a conjecture and prove it). Pre-service teachers in the heuristic worked examples condition received additional worked examples, in which a fictitious student proves a conjecture using mathematical theories and operations, principles of mathematical proof and heuristic strategies. The teachers' performance on a post-test showed that heuristic worked examples (compared to problem-solving) could effectively be used to facilitate mathematical argumentation.

Furthermore, worked examples can also be effective when they include *errors* (Große & Renkl, 2004). For example, Kopp et al. (2008) conducted an experimental study with a 2x2 factorial design with two different kinds of worked examples (erroneous vs.

not erroneous) and feedback (elaborated vs. knowledge of correct result) to facilitate diagnostic knowledge of medical students ($N = 158$). After completing a pre-test on diagnostic knowledge, students worked through the worked examples. The worked examples illustrated a fictitious student conducting a diagnostic process in a clinical situation. In the group without errors, the student comes to the correct solution. In the erroneous condition, the student commits several errors. Results of a post-test on students' diagnostic skills showed that erroneous examples were more effective regarding the facilitation of diagnostic knowledge than correct examples, especially when they were combined with elaborated feedback. Kopp et al. (2009) partially replicated the results. Furthermore, while worked examples are in general considered to be especially promising in early phases of skill acquisition, empirical evidence from Große and Renkl (2007) in the context of a well-structured domain (solving probability problems) suggests that erroneous worked examples can be effective when learners already hold some prior knowledge of the subject matter.

4.3.2 Scripts to support Students' Enactment of Fallacy Recognition

Worked examples push the learner in a rather passive role, as she is only supposed to study the example (rather than to solve problems on her own). When being exposed to an external script, in contrast, learners are asked to adopt a much more active role, since scripts are intended to induce certain activities (that may relate to the underlying principles of the solution paths resembled by worked examples; Renkl, 2014). Therefore, they can be used to overcome potential learners' passivity by inducing cognitive processes through which the learner may apply principles of a solution process.

In psychology, the term *script* has two different notions: *Internal scripts* refer to knowledge structures that someone already holds. *External scripts* refer to knowledge structures provided externally, for example through a learning environment. Chapter 4.3.2.1 introduces the main assumptions of internal and external scripts. In general, scripts are knowledge structures that refer to procedural knowledge, that is, *how* knowledge is applied to solve a certain task or a problem (Anderson, 1996). A learner might already possess such knowledge structures (internal scripts) that are relevant for a given problem. However, these internal scripts might not be ideally configured regarding the performance of certain tasks that must be carried out to solve the problem in an ideal way (Fischer, Kollar, Stegmann, & Wecker, 2013). Educational environments can support students in their development of such knowledge structures by providing them with external support, which also has been termed *external scripts* in related research (e.g., Kollar, Fischer, & Slotta, 2007). In this context, however, the term does not refer to memory structures, but rather to a specific kind of an externally provided support (i.e., a scaffold). Chapter 4.3.2.2 outlines the Script Theory of Guidance which provides a theoretical account that explains how external scripts work. Chapter 4.3.2.3 explains how

external scripts can be designed and provides empirical evidence on the effectiveness of such scripts.

4.3.2.1 Definition of Internal and External Scripts

The term *script* was coined by Schank and Abelson (1977). It has two different notions, one regarding its use in cognitive psychology (*internal scripts*), and one that has its origins in educational psychology (*external scripts*).

With reference to cognitive psychology, Schank and Abelson (1977) proposed that humans organize procedural knowledge mentally in an episodic manner (i.e., based on experiences rather than on semantic categories) through cognitive schemata (scripts). Scripts provide procedural knowledge that helps a person to understand and act in a specific situation (i.e., move from one problem state to the next; de Jong & Ferguson-Hessler, 1996). Kollar et al. (2007) termed such internal knowledge structures *internal scripts.* A prominent example is the *restaurant script*, which a person might establish through repeated experience (i.e., several restaurant visits). Due to this script, a person who is about to go out for dinner has a clear idea of the sequences that will be part of such an evening. For example, the person will wait to be seated, then pick a table, receive the menu, choose a dish and so on. An individual might know this procedure from the experiences of former visits of restaurants before even entering the restaurant and can therefore act accordingly. Likewise, an experienced social worker might have established a script for a social work decision-making context such as a professional meeting, in which arguments are exchanged and evaluated to arrive at a conclusion regarding the potentially best intervention for a client. Such a script might, for example, entail a problem conceptualization, evidence generation and evaluation (resp. evaluation of arguments), problem explanation, prediction of changes, ethical assessment, formulation of goals and the decision for an intervention (Geiser, 2015; Staub-Bernasconi, 2018).

However, individuals may lack a certain internal script or hold one which is not ideal for a given problem at hand (Kollar et al., 2007). For example, a first-year social work student will probably not hold an ideal internal script for a social work decision-making context, given he never participated in such contexts. According to the Script Theory of Guidance (Fischer et al., 2013; chapter 4.3.2.2), one way to compensate lacking or dysfunctional internal scripts is to provide learners with *external scripts.* External scripts are instructional means that help learners acquire internal scripts. In the case of a social work decision-making context, this might mean to provide the learner with more specific knowledge that relates to the steps of a social work decision-making context script. For evidence evaluation (resp. argument evaluation), such specific knowledge could refer to the question how to build an argument (conclusion and claim/s) and criteria for good argumentation (Damer, 2011). Hence, a goal of external scripts is often to support individual knowledge acquisition (Kollar, Fischer, & Hesse, 2006). This knowledge can refer to skills (procedural knowledge) and/or declarative knowledge which is supposed to be acquired more easily through the application of skills. An external script has two important functions: it (1) prompts the learner to carry out certain activities and it (2) provides information on when to carry out which activity (Kollar et al., 2007). It represents sequences of practices (e.g., textual or graphical) in a way that is supposed to be ideal to perform necessary activities in a given situation (e.g., a problem-solving process). That way, external scripts guide the learner through cognitive activities. Herein lies the difference (but also the complementary nature) of external scripts to worked examples. While worked examples usually show the learner the solution process and the actual solution, external scripts prompt the learner to perform the solution process. The Script Theory of Guidance provides an elaborated theoretical account regarding the interplay of internal and external scripts, thereby

pointing out how guidance through external scripts should ideally look like. The Script Theory of Guidance will be explained in the following chapter.

4.3.2.2 Script Theory of Guidance

The Script Theory of Guidance (Fischer et al., 2013) intends to provide a theoretical basis for the script-based facilitation of collaboration skills. However, in the context of this thesis it is applied to individual learning contexts (i.e., the facilitation of fallacy recognition skills by individual learners). An important theoretical assumption of the Script Theory of Guidance is that cognitive knowledge structures are not stable (Schank & Abelson, 1977) but rather dynamic (Schank, 1999). A second important assumption is that individuals can increase their knowledge and skills by participating in activities that are theoretically beyond what they are currently capable of when they are supported by more knowledgeable others (e.g., by scaffolds). This idea relates to the so-called Zone of Proximal Development (ZPD; Vygotsky, 1978). The ZPD "is the distance between the actual developmental level as determined by independent problem-solving and the level of potential development as determined through problem-solving under adult guidance or in collaboration with more capable peers." (Vygotsky, 1978, p. 86). In other words, there are knowledge-related activities a learner can perform without guidance, activities a learner can perform with guidance, and activities a learner cannot perform, even with guidance. For example, a social work student might currently be unable to evaluate an argument against criteria for good argumentation (Damer, 2011) without support, but he might be able to perform such an evaluation with additional guidance. However, he might be unable to perform the process of EBP, even with additional guidance, because he currently lacks important related prerequisites such as fallacy recognition skills. *Guidance* refers to an external source (e.g., a teacher or an external script) that can provide knowledge that the learner is yet lacking. The goal of guidance is to induce a

reconfiguration of a learner's internal script in such a way, that the reconfigured internal script allows the learner to perform similar future tasks on his own. The Script Theory of Guidance states seven principles concerning this process. However, only four of these are relevant in the context of this doctoral thesis. These four principles are: (1) internal script guidance principle, (2) internal script induction principle, (3) external script guidance principle and (4) optimal external scripting level principle.

According to the *internal script guidance principle*, a learner understands and performs a task in accordance with her respective internal script. This internal script is configured and reconfigured dynamically (i.e., it is not stable and can be changed) and consists of three[5] different components – play, scene and scriptlet – that are based on dynamic memory theory (Schank, 1999). These three components have an hierarchical order (in the sense that a higher order component entails lower order components, which in turn means that lower order components have to be learned before higher order components can be applied properly) with *play* being the highest and *scriptlet* the lowest component. Considering the goals of this thesis, a fourth component is proposed. This component is labeled *cognitive operations* and it is hierarchically located under the scriptlet component. Thus, the four components are (in hierarchical order): (a) Play, (b) scene, (c) scriptlet and (d) cognitive operations.

The *play* component entails and organizes knowledge about the task at hand (e.g., a decision-making process in social work) and the sequences of scenes that are part of such a task (e.g., critical appraisal of relevant evidence presented within a discourse of the decision-making process).

[5] The Script Theory of Guidance originally states *role* as a fourth component. This refers to different roles a learner may have during collaboration processes such as a commenter or observer. However, the role component is not relevant for the goals of this thesis.

The *scene* component includes knowledge about the situations that are part of the *play* component, such as the conceptualization of a problem, the collection of relevant data or the critical appraisal of the evidence generated from that data within a decision-making process (the play component).

The *scriptlet* component entails knowledge about tasks that are part of a *scene*, such as evaluating an argument as part of the critical appraisal scene.

Finally, the *cognitive operations* component constitutes knowledge that guides the learner through a task that is associated with a scriptlet component, such as identifying conclusions and claims as a task that is part of the evaluation of an argument (the scriptlet component).

According to the *internal script induction principle*, learners reconfigure available internal script components when they encounter a new and unfamiliar task. If the learner perceives a new configuration as successful, repeated practice might contribute to the development of new higher-order components.

The *external script guidance principle* states that external scripts enable learners to perform tasks that would be beyond their capabilities without external support (cf. ZPD; Vygotsky, 1978). They do this by either facilitating the use of internal script components that are considered ideal for a given task or by inhibiting the use of internal script components that would be rather dysfunctional for the problem at hand. In the context of argument evaluation in a social work decision-making context with respect to the recognition of fallacies, an internal script that relates to the identification of a conclusion and (a) claim/s as well as to the evaluation of the conclusion and the claim/s against the five criteria for good argumentation (Damer, 2011) might be considered ideal.

In contrast, for example, a script that induces argument evaluation by directly assessing the conclusion (i.e., without considering related claims and/or criteria for good argumentation) might be considered suboptimal.

The *optimal external scripting level principle* states that an external script should always target the highest possible script component level. This means that if a learner has, for example, already developed and organized script components at the *scriptlet* level in an ideal way, but not at the *scene* level, then, an external script should target the *scene* level instead of any other level to provide the learner with an optimal ZPD. If it would target the *scriptlet* level, it aims at tasks the learner is already capable of performing without guidance. The external script is thus likely to include unnecessary information. This in turn would raise extraneous load (however, the concept of cognitive load is not part of the original Script Theory of Guidance; it is integrated here to support coherence and comprehensibility), since the learner would have to process all this information, which hinders knowledge acquisition. However, if the external script would target the *play* level, the external script would aim at tasks the learner cannot perform, even with guidance. In this example, only an external script that targets the *scene* level would provide the learner with the optimal ZPD.

An idea that is related to the optimal external scripting level principle is *overscripting* (Dillenbourg, 2002). Overscripting might occur when too much information (i.e., too much scaffolding) is provided. In this case, cognitive load is raised due to the necessity for the learner to process (i.e., understand, memorize and apply) the script. Since understanding how to process the script is not directly related to knowledge and skill acquisition regarding the subject matter, overscripted external scripts may impair learning processes.

Empirical evidence regarding overscripting comes from a study by Mäkitalo, Weinberger, Häkkinen, Järvelä, and Fischer (2005) in which they investigated effects of a cooperation script with regard to collaborative learning. 48 educational science students were randomly grouped into triads. These triads were then randomly assigned to either a group with the cooperation script or a group without the cooperation script. Each students of a triad were placed in a separate room. The students had to read a text about a theory. Afterwards, they had to analyze three problems described in a vignette by applying concepts of the theory and discuss it in an online environment. Students in the condition without the script received no support to solve the cases. In the script condition, students received prompts to find relevant problem information and apply the theoretical concepts to this information. The prompts were considered a scaffold on the scriptlet level. A post-test with a problem case measured individual learning outcomes. Results for participants in the unscripted condition were significantly higher than for those in the scripted condition. This might indicate that students already had adequate scriptlet components to solve the problem case. Therefore, the scaffolds might have induced unnecessary extraneous load.

4.3.2.3 Design and Effects of External Scripts

External scripts are particularly used in research on computer-supported collaborative learning. There, they are either applied to induce activities that learners need to perform to solve a certain problem, or they are used to guide interactive processes between collaborators (Kollar et al., 2006). In collaborative learning, external scripts include five important aspects, namely a specific (1) *objective* (e.g., induce discourse), (2) *learning activities* (e.g., questioning, explaining), (3) *sequencing* (i.e., specifying when leaners should perform which activity), (4) *role distribution* (e.g., explainer, commentator) and

(5) *type of representation* (how the external script is presented to the learner; Kollar et al., 2006).

The first four aspects can be considered rather self-explanatory, however, regarding the *type of representation*, several aspects must be considered. External scripts can be presented in different ways, such as text (King, 1998), graphics (Pfister & Mühlpfordt, 2002) or spoken words (Palinscar & Brown, 1984). Their design is supposed to take learners' internal scripts into account (Carmien, Kollar, Fischer, & Fischer, 2007; Fischer et al., 2013; Kollar et al., 2007) to provide them with an optimal ZPD. Furthermore, external scripts vary in their degree of freedom or *coercion*. Some scripts might grant the learner a lot of autonomy regarding her approach to solve a particular problem (e.g., ask a learner to evaluate an argument without further instructions), while other scripts can be very strict and force the learner to perform specific activities in a specific order (e.g., ask the learner to identify the conclusion of an argument, then to identify its claims, then to evaluate identified claims and conclusions against the structure criteria, and so on; Kollar et al., 2006). Finding the right degree of structuredness is important to grant learners an optimal balance between external support and self-regulation (Wecker, Kollar, Fischer, & Prechtl, 2010). As stated by the optimal external scripting level principle, an external script should always target the highest possible script component level (Fischer et al., 2013). Yet, the question to which degree external scripts should be specified in order to prevent effects related to overscripting is still underinvestigated (Noroozi, Weinberger, Biemans, Mulder, & Chizari, 2012).

Empirical evidence on the effectiveness of external scripts comes mainly from research on collaborative learning. It suggests that external scripts can have positive effects on the acquisition of skills in collaborative settings. Empirical studies show that

this accounts also for argumentation skills. For example, Stegmann, Weinberger, and Fischer (2007) conducted an experimental study in which they applied an external script to facilitate the quality of single arguments as well as argumentation sequences among students ($N = 120$) from educational science. The script for single arguments provided learners with text boxes for a claim, grounds and qualifications. That way, learners were prompted to construct an argument in the respective way. The script for argument sequences was intended to facilitate sequences of argument-counterargument integration and labeled arguments accordingly. Learners in the control condition received no additional support. The script for single arguments yielded a significant effect that strongly affected the acquisition of knowledge on the construction of single arguments ($\eta^2 = .22$) and the script for argument sequences yielded a significant effect that strongly affected the acquisition of knowledge regarding the construction of argumentation sequences ($\eta^2 = .32$), as indicated by the post-test.

Scripts can have positive effects on domain-general (e.g., argumentation skills) as well as domain specific (e.g., diagnostic competencies in medicine) skills (see Fischer, Chinn, Engelmann, & Osborne, 2018 for an elaborated account on domain-general and domain-specific knowledge). Fallacy recognition skills relate to argumentation (Damer, 2011), however, fallacies are strongly context-bound (Walton, 1995). Therefore, it seems reasonable to regard fallacy recognition skills in social work decision-making contexts more as a form of cross-domain skills that entail aspects of both classical dichotomously distinguished knowledge forms (Hetmanek, Engelmann, Opitz, & Fischer, 2018). Hence, potential effects of external scripts on both, domain-specific as well as domain-general skills are of interest in the context of this doctoral thesis. In their meta-analysis, Vogel, Wecker, Kollar, and Fischer (2017) summarize empirical research to investigate effects

of collaboration scripts on domain-specific knowledge as well as on collaboration skills (which can be considered rather domain-general knowledge). Twenty-two articles met their inclusion criteria. These articles include 24 studies from various countries and disciplines such as natural sciences, social sciences and technical disciplines with a total of 34 comparisons. They found external scripts to be effective compared to unstructured collaborative learning regarding the acquisition of *domain-specific knowledge*, however, with a small effect size ($d = 0.20$). On the other hand, the overall effect size for *collaboration skills* was substantial ($d = 0.95$). The subject discipline was identified as a significant moderator for collaboration skills. External scripts were found to be especially effective in that regard, when they are combined with content-specific scaffolds (e.g., a worked example).

Empirical evidence regarding effects of external scripts specifically on critical thinking skills is rather limited, but results from respective studies suggest that external scripts can enhance critical thinking skills. For example, Schellens, van Keer, Wever, and Valcke (2009) investigated effects of a collaboration script (with vs. without) on skills associated with critical thinking (e.g. problem identification, problem exploration) of university students ($N = 35$). Students had to discuss possibilities and constraints of e-learning as a solution for educational problems in online discussions. The experimental group was provided with a checklist that resembled the various critical thinking skills. Students had to label their contributions accordingly, thereby indicating the contribution to the discussion progress. Answers were coded with respect to critical thinking categories such as outside knowledge or critical assessment. Students in the scripted condition showed significantly more positive and less negative indicators of critical thinking than students in the unscripted condition. In another study regarding the effects

of external scripts on critical thinking, Y. H. Lee (2015) developed a computer-supported external collaboration script based on the Script Theory of Guidance to facilitate critical thinking skills of educational psychology students ($N = 78$) from a university in Taiwan. The experimental group received an external script that specified how participants should interact with regards to the mutual work on an article on wetland restauration (pose and discuss questions based on headlines, read article individually, answer questions and take notes, check if questions can be answered and complete notes), whereas the control group did not receive further guidance. Students in the experimental condition outperformed students in the control condition in a post-test (multiple choice) regarding their overall scientific literacy skills ($\eta^2 = .06$). Y. H. Lee (2018) used the same script successfully in another experiment with two groups (with script vs. without script) to foster scientific reading skills of Taiwan psychology students ($N = 85$). No significant differences were found in a post-test with multiple choice questions regarding scientific reading skills. However, students in the scripted group scored significantly higher than the unscripted group in a follow-up test (1 week; $d = 0.48$).

Taken together, empirical evidence suggests that external scripts can facilitate skills that are related to critical thinking. However, scripts have not been applied in research on the facilitation of such skills within the domain of social work. Therefore, it is yet an open question to which extent they can facilitate such skills in social work contexts.

4.4 Summary

This chapter introduced instructional approaches and scaffolds regarding the facilitation of Evidence-based Practice and fallacy recognition skills. Chapter 4.1 outlined

instructional approaches, specifically, more *teacher*- and more *student-centered* ones. Teacher-centered approaches (chapter 4.1.1) structure learners' learning processes to a greater extent, while student-centered approaches (chapter 4.1.2) grant learners more freedom in that regard.

Chapter 4.2 introduced *Case-based Reasoning* as an example for a rather student-centered instructional approach in which learners learn with meaningful and authentic problem-based cases. Such a learning process may facilitate experiences that can be cognitively stored. If a learner encounters a scenario such as a problem-solving situation that is similar to such a cognitively experience, recalling the experience that is most appropriate in providing guidance on how to approach the similar problem may support her in problem-solving processes. Empirical evidence suggests that Case-based Reasoning is a promising means to facilitate skills that are closely related to fallacy recognition skills (e.g., critical thinking). The effectiveness of Case-based Reasoning can be further enhanced through scaffolds such as worked examples and external scripts.

Worked examples (chapter 4.3.1) are instructional means that provide a problem, a solution process and an actual solution. They have been shown to be especially effective in well-structured domains (e.g. mathematics), but empirical evidence suggests that they can also be effectively applied in ill-structured domains (e.g., medicine). Therefore, it is reasonable to assume that worked examples can also be effectively applied in an ill-structured domain such as social work. Empirical evidence suggests that erroneous worked examples might be especially promising in ill-structured domains (chapter 4.3.1.3). Furthermore, they can also be beneficial for learners that already hold some prior knowledge. Effects (chapter 4.3.1.1) of worked examples can be explained with Cognitive Load Theory (chapter 4.3.1.2). It was argued that worked examples reduce cognitive load in the learner's working memory during a learning process by pointing the

attention of the learner to aspects that are relevant for learning. That way, schema acquisition becomes more likely. However, worked examples can be considered a rather passive instrument (i.e., learners look at them without necessarily taking further action). Scripts are a more active means in that they intend to induce further activities.

Scripts are knowledge structures that refer to the question how activities are carried out (chapter 4.3.2.1). They can be internal (i.e., knowledge structures that someone already holds) or external (i.e., knowledge structures that are externally provided to induce and guide certain activities). It was argued that scripts prompt the learner to apply the underlying principles of the worked examples. Chapter 4.3.2.2 outlined the Script Theory of Guidance as a theoretical account for an explanation regarding the interplay of internal and external scripts. It was argued that internal scripts entail four different components with a hierarchical order. When designing an external script, the configuration of learners' respective internal scripts should be considered to provide them with an optimal Zone of Proximal Development.

5. Implications for the Empirical Studies of this Thesis

This chapter bridges the theoretical considerations (Part I) with the empirical studies (Part II). Chapter 5.1 draws on specific parts of the theoretical considerations and presents respective implications for the first empirical study, a systematic review regarding the question how to teach the process of Evidence-based Practice as well as the use of Empirically Supported Treatments in social work (chapter 6). Chapter 5.2 presents conclusions that follow from the theoretical introduction for the second empirical study, an experimental study about effects of worked examples and an external script on social work students' fallacy recognition skills in a case-based learning environment (chapter 7).

5.1 Implications for Empirical Study 1

Chapter 2 introduced Evidence-based Practice and its two approaches that are most prevalent in social work, namely the process of Evidence-based Practice and Empirically Supported Treatments. It was argued that, despite considerable amounts of criticism, it is an important endeavor to implement and use these approaches in social work practice. It was illustrated that empirical evidence suggests that their implementation lags and that insufficient education of the approaches in the field of social work is one reason for this shortcoming. Thus, the goal of the first study is to find out which instructional approaches (chapter 4.1) are most promising to teach the use of Evidence-based Practice approaches to social work students and social workers. A suitable methodological approach to tackle this question is a systematic review. It provides means to systematically search for and evaluate empirical studies that investigated educational means with the goal to facilitate

knowledge about and the use of Evidence-based Practice approaches amongst social work students and social workers (Higgins & Green, 2011; Liberati et al., 2009).

5.2 Implications for Empirical Study 2

In chapter 3 it was argued that skills that are related to argument evaluation are considered important prerequisites for Evidence-based Practice and particularly for the process of Evidence-based Practice. In this context, argument evaluation was conceptualized as fallacy recognition (chapter 3.2). Subsequently, skills that refer to fallacy recognition were specified as the skill to *identify* fallacies in social work decision-making contexts, as well as the skill to *explain* why a fallacious argument is fallacious. Both skills were summarized under the term *fallacy recognition skills*. The argument model of Damer (2011) was introduced as a respective theoretical account as well as the related five criteria for good argumentation (chapter 3.1 and 3.2). A fallacy was conceptualized as an argument that violates one or more of these criteria for good argumentation (chapter 3.2). It was outlined why it is important to investigate the most effective ways regarding the facilitation of fallacy recognition skills in social work. Towards that end, chapter 4.2 introduced *Case-based Reasoning* as a rather student-centered instructional approach that seems to be promising in facilitating social work students' fallacy recognition skills. It was explained that student-centered approaches can be particularly effective, when learners already hold some prior knowledge of the subject matter. In this regard, it is reasonable to assume that advanced (e.g., sixth semester) social work students may already hold some prior knowledge that is relevant for argument evaluation. For example, social work students in Bavaria (Germany) complete a field practice unit (20 weeks) in their fifth semester. There, they participate directly in practical contexts that will most likely provide them with experiences in

formulating professional arguments to arrive at a reasonable suggestion for the potentially best intervention for a given client. Accordingly, a case-based learning environment will be used for the second empirical study. The effectiveness of Case-based Reasoning might be enhanced through *scaffolds* (i.e., a kind of support that guides a learner through certain activities), namely *worked examples* and *external scripts*. Worked examples (chapter 4.3.1) were introduced as a scaffold that show the learner a problem, a solution process and the solution. That way, they point the learner's attention to aspects that are relevant for learning which may reduce extraneous load in a learner's working memory, thereby making schema acquisition more likely. Erroneous worked examples were shown to be potentially effective in ill-structured domains and for learners who already hold some prior knowledge. Social work can be considered an ill-structured domain. Furthermore, it can be speculated that advanced social work students hold at least some prior knowledge regarding argument evaluation (e.g., due to experiences in their field practice). These insights must be taken into account when designing the worked examples. While worked examples can be considered a rather passive scaffold, scripts (chapter 4.3.2) were introduced as a more active one. External scripts can induce activities (such as argument evaluation tasks, chapter 3.1) and guide the learner in applying their related principles (e.g., criteria for good argumentation, chapter 3.2). That way, external scripts can help a learner to approach a problem-solving process in a potentially more ideal way, compared to the way in which she would tackle a problem based on her current internal scripts (given that these internal scripts are indeed less optimal than the provided external scripts). Accordingly, the goal of the second empirical study is to investigate the effectiveness of worked examples as well as of external scripts on social work students' fallacy recognition skills in social work decision-making contexts on the basis of Damer's conceptualizations of arguments and

fallacies within a case-based learning environment. A suitable methodological approach regarding this question is an experimental study, since it allows to compare different conditions and causal inferences due to randomization and potentially high internal validity (Shadish, Cook, & Campbell, 2002).

Part II

Empirical Studies[6]

[6] According to the study regulations (§ 17), Part II contains one article published in (Study 1) and one article submitted to (Study 2) international peer-reviewed journals.

6. Study 1 – How to Teach Evidence-based Practice in Social Work: A Systematic Review[7]

This article is published in the Journal of Research on Social Work Practice by SAGE and can be accessed via the following URL:

https://doi.org/10.1177/1049731519852150

[7] This article was co-authored by Ingo Kollar, Eileen Gambrill, Christian Ghanem and Sabine Pankofer.

Research Article

Research on Social Work Practice
1-21

DOI: 10.1177/1049731519852150
journals.sagepub.com/home/rsw
SAGE

How to Teach Evidence-Based Practice in Social Work: A Systematic Review

Florian Spensberger[1], Ingo Kollar[2], Eileen Gambrill[3], Christian Ghanem[4], and Sabine Pankofer[5]

Abstract

Purpose: This article presents a systematic review of research regarding how best to educate social work students and practitioners concerning of the process of evidence-based practice and/or the application of empirically supported treatments (ESTs). **Method:** We conducted a systematic review with a narrative synthesis, largely following the *Cochrane Handbook of Systematic Reviews for Interventions* and *PRISMA* reporting guidelines for systematic reviews and meta-analyses. **Results:** Twenty-seven studies met our eligibility criteria. These consisted mostly of uncontrolled designs and their measures relied mainly on learners' self-perception regarding acquisition of declarative and procedural knowledge, motivation, and satisfaction. Reports were mostly positive (88.7%). **Conclusions:** Research regarding the education of social work students and practitioners about the process of evidence-based practice as well as ESTs is limited. Further investigation is needed concerning the effectiveness of specific teaching methods using controlled designs and more rigorous outcome measures including observation of practice in real-life situations and/or in role-plays.

Keywords

evidence-based practice, empirically supported treatments, education, social work, systematic review

It has been argued that the process of evidence-based practice (EBP) will contribute to making informed decisions that help clients attain valued outcomes (Emparanza, Cabello, & Burls, 2015; Gambrill, 2006; Sackett, Richardson, Rosenberg, & Haynes, 1997; Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996). In EBP, compared to authority-based approaches (Gambrill, 1999), currently available research related to particular clients is sought as well as information about client circumstances and characteristics including their preferences and values, and clinical expertise is drawn on to integrate all information. Uncertainty and ignorance as well as knowledge is shared among professionals and clients. In authority-based approaches, criteria such as consensus and tradition are relied on in making decisions. Ever since EBP was promoted in social work (Gambrill, 1999), it has sparked interest. Two different approaches emerged: (1) the process of EBP and (2) empirically supported treatments (ESTs; promotion of specific interventions) also referred to as evidence-based interventions (EBIs) or evidence-based practices (EBPs). In the following, we will refer to all of these terms as ESTs for easier readability. Since there are different views of (Rubin & Parrish, 2007) and misconceptions about EBP (Gibbs & Gambrill, 2002), both approaches are addressed in this review. To date, little is known about how ESTs and/or the process of EBP are typically taught in social work education (or if they are). Thus, the aim of this article is to systematically describe the state of research on how to best teach the process of EBP and/or ESTs to social work students and practitioners as well as with regard to its quality.

Evidence-Based Practice: Two Different Approaches

There are two main different understandings of EBP. One is the process of EBP as described in original sources in medicine designed to help practitioners make informed decisions (Haynes, Devereaux, & Guyatt, 2002; Sackett et al., 1996; Straus, Richardson, Glasziou, & Haynes, 2011). A second (ESTs) refers to interventions claimed to be effective by some individual or organization. Both approaches are briefly discussed next (see Thyer & Myers, 2011, for an elaborated distinction).

The process of EBP. The term "evidence-based medicine (EBM)" was coined by Guyatt (1991; see Sur & Dahm, 2011, for a

[1] Department of Psychology, Ludwig-Maximilians-Universität München, Munich, Bavaria, Germany
[2] Faculty for Philosophy and Social Sciences, Universität Augsburg, Augsburg, Germany
[3] School of Social Welfare, University of California Berkeley, Berkeley, CA, USA
[4] Department of Social Work, Katholische Stiftungshochschule München, Benediktbeuern, Munich, Germany
[5] Department of Social Work, Katholische Stiftungshochschule München, Munich, Germany

Corresponding Author:
Florian Spensberger, Department of Psychology, Ludwig-Maximilians-Universität München, Leopoldstraße 44, Munich 80802, Germany.
Email: florian.spensberger@psy.lmu.de

description of the history of EBM). In the process of EBP, clinical expertise is drawn on to integrate relevant research findings, and information about the clients' unique circumstances and characteristics including their values and preferences, and hoped-for outcomes in order to arrive at informed decisions. This process involves "the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual [clients]" (Sackett et al., 1997, p. 2; see also Sackett et al., 1996). Clinical expertise includes basic skills of clinical practice, including relationship skills and the practitioner's individual experience (Haynes et al., 2002). The process of EBP includes five steps as described in original sources (Sackett et al., 1996; Straus et al., 2011).

1. Converting information needs related to practice decisions into well-structured questions.
2. Tracking down, with maximum efficiency, the best evidence with which to answer those questions.
3. Critically appraising that evidence for its validity, impact (size of effect), and applicability (usefulness in practice).
4. Integrating this critical appraisal with clinical expertise and with a client's unique circumstances and characteristics including their values and preferences and making a decision together with the client.
5. Evaluating the effectiveness and efficiency in carrying out Steps 1-4 and seeking ways to improve them in the future.

This approach requires a search for knowledge as well as for ignorance such as lack of relevant research concerning a problem (Gambrill, 2019). Results are shared with clients to enable informed decisions that are most likely to result in hoped-for-outcomes for clients.

ESTs. The term "empirically supported treatments" (other terms include EBPs, empirically tested interventions, and EBIs) refers to manualized interventions (e.g., cognitive behavioral therapy, motivational interviewing) deemed to be "empirically supported" based on related research (Thyer & Myers, 2011). For example, the American Psychological Association 2005 Presidential Task Force on Evidence-Based Practice suggested criteria for categories such as "well-established" (at least two good group design studies or a large series of single case design studies, study conducted with treatment manual, clearly specified sample characteristics) and "probably efficacious" (e.g., two studies showing that a treatment is more effective than a waiting-list control group, Task Force, 1993).

Implementation of EBP in Social Work

Even though EBP has become an intensively discussed topic within social work, its implementation in social work practice still lacks behind. With regard to EBP as a process, Pope, Rollins, Chaumba, & Risker (2011) found in a survey of social workers ($n = 200$) in the United States that, although 83% agreed or somewhat agreed to be familiar with social work databases, only 56% agreed or somewhat agreed that they used relevant research to answer clinical questions (range: agreed, somewhat agreed, undecided, somewhat disagree, disagree). In a survey by Parrish and Rubin (2012) of 688 social workers carried out in Texas (a 21% response rate), it was found that few social workers indicated on a 5-point Likert-type scale that they "often" or "very often" use the Internet to search for the best evidence to "guide practice decisions" (32.8%), "read about research evidence to guide practice decisions" (37.8%), "inform clients of the degree of research evidence supporting alternative intervention options" (25.6%), and "engage in all steps of the EBP process" (15.1%). With regard to ESTs, Morago (2010) reported that 42.6% of 155 social workers and social care professionals indicated the level of implementation of ESTs in their respective agency as "very poor" and 40% as "modest" in a survey conducted in the UK (range: very poor, modest, good, excellent).

Research in a variety of professions has shown that implementation of EBP is difficult due to numerous barriers (e.g., Gray, Joy, Plath, & Webb, 2012; Scurlock-Evans & Upton, 2015). Skill and knowledge may be lacking. There may be insufficient preparation to use EBP (Teater & Chonody, 2018), unsound training (Bellamy, Bledsoe, & Traube, 2006), negative attitudes toward EBP (Murphy & McDonald, 2004), and diverse views of EBP (Rubin & Parrish, 2007). Therefore, social workers may be ill-prepared to use either ESTs and/or the process of EBP. It is important to identify effective educational interventions (EIs) to help students and practitioners to acquire and use related knowledge and skills if these enhance success in helping clients.

Systematic Reviews on EBP Education in Other Areas

The production of systematic reviews has greatly increased over the past decades. Yet, many reviews have been criticized as flawed (Ioannidis, 2016). There are a number of systematic reviews concerning the process of EBP in medicine. Aglen (2016) conducted a systematic review with 39 articles to provide an overview of strategies used to teach the process of EBP to nursing students at the bachelor level. Most studies ($n = 31$) used a qualitative, descriptive design and a formative evaluation; the focus was on students' satisfaction with the EIs. Aglen (2016) concluded that a key issue in teaching the process of EBP was that nursing students do not see how research findings can contribute to their practice. She argued that teaching critical thinking and an emphasis on clinical problems are important aspects in teaching the process of EBP. Dizon, Grimmer-Somers, and Kumar (2012) reviewed six studies (four randomized controlled trials and two studies that used a single group pre–posttest design) regarding the effectiveness of training programs for the process of EBP to improve the knowledge, skills, attitudes, and behavior of allied health professionals (e.g., physiotherapists, speech pathologists). They concluded that training had significantly positive effects on all of the aforementioned constructs. Kyriakoulis et al. (2016) reviewed

20 studies (4 controlled and 16 uncontrolled trials) to find the best strategies and methods to teach the process of EBP to undergraduate health students. They concluded that a multifaceted approach that entails a combination of methods like lectures, computer sessions, small group discussions, journal clubs, and assignments was more likely to improve learners' EBP knowledge and motivation than interventions offering only one of these methods or no intervention. Patelarou et al. (2017) reviewed 20 studies (3 controlled trials and 17 uncontrolled) to find the best teaching strategies and methods used to teach the process of EBP to health professionals and found that an increase in EBP competencies and attitudes was reported in nine of them. These authors also recommend a multifaceted teaching approach. In addition, they reported that online EIs are effective in enhancing practitioners' claimed motivation to use the process of EBP. They also reported that online interventions were not effective in achieving changes in self-reports of behavior (however, the review also offered some data that support the effectiveness of online interventions regarding EBP behavior).

All of these reviews were conducted in fields different from social work. We could not find a review regarding teaching EBP in social work—neither with respect to the process of EBP nor the application of ESTs. Thus, a systematic review of research on how to teach EBP in social work is lacking.

How to Teach EBP: Instructional Approaches and Knowledge Application

The question how to best teach the process of EBP and/or ESTs can be tackled from different perspectives. One research community that is particularly concerned with the teaching of complex skills is the Learning Sciences community (e.g., Fischer, Hmelo-Silver, Goldman, & Reimann, 2018; Sawyer, 2009; see Hoadley & van Heneghan, 2012, for a brief history of the Learning Sciences and their implications for instructional designs). To categorize different teaching approaches, Learning Sciences research has repeatedly differentiated between "teacher-centered" approaches on the one hand (approaches that view the teacher as the main instance regarding what and how to learn in the classroom) and more "learner- or student-centered" approaches on the other hand (approaches that provide learners with more freedom regarding how to structure their learning process). It is argued that these concepts provide a useful analytical segregation for empirical research on EIs and its potential implications (see Hmelo-Silver, Duncan, & Chinn, 2007; Kirschner, Sweller, & Clark, 2006; Sweller, Kirschner, & Clark, 2007, for a critical discussion of this dichotomy). Direct instruction (DI, e.g., Slavin, 2018) is an example for the teacher-centered approach. Problem-based learning (e.g., Hmelo-Silver, 2004) is an example of the student-centered approach. In the following, we describe the two approaches and their respective examples in more detail.

Teacher-centered instructional approaches. The basic idea of teacher-centered approaches is to have a teacher to support student learning by providing information that explains concepts and procedures (Kirschner et al., 2006) optimally in a way that "fits" the human cognitive architecture (especially not to overstrain leaners' working memory capacity; Sweller, Ayres, & Kalyuga, 2011). DI is an example of this approach. Slavin (2018) suggests seven steps to apply this approach in an ideal way: (1) define learning goals and provide a syllabus, (2) activate prior knowledge, (3) present new subject matter in a structured and efficient way, (4) use comprehension checks like questions, (5) let learners apply previously presented knowledge, (6) induce further elaboration (e.g., homework), and (7) assess performance and give feedback. In a meta-meta-analysis, Hattie (2009) reported an average effect size of $d = .59$ for DI in comparison with other traditional teaching approaches. If delivered correctly, DI helps to avoid exposing too much load on the leaners' working memory (Sweller et al., 2011). However, knowledge acquired through teacher-centered approaches often remains inert, that is, it is often difficult for learners to use this knowledge for problem-solving. One possible explanation is that the acquisition of knowledge is context-bound. Transferring that knowledge to a situation that is very different from the situation in which it was acquired can be very difficult (Barnett & Ceci, 2002). Situated (e.g., student-centered) approaches have been developed to tackle this issue (Renkl, Mandl, & Gruber, 1996).

Student-centered instructional approaches. In student-centered instructional approaches, learners are granted a more active role. This is achieved by presenting learners more complex and practical problems that they are supposed to solve either alone or in groups but optimally guided by a teacher or tutor. One example is problem-based learning (PBL; Barrows & Tamblyn, 1980; Hmelo-Silver, 2004). In PBL, after the presentation of a problem, students discuss possible explanations for it. Discussing the problem before receiving any further instructions is important to activate and evaluate prior knowledge and discover knowledge gaps that should trigger interest and motivation to find out more about the problem (Loyens & Rikers, 2011). In PBL, students learn by solving complex real-world problems and reflect on their experiences guided by a teacher or a tutor (Hmelo-Silver et al., 2007). In Hattie's (2009) meta-meta-analysis, the average effect of PBL on student achievement compared to more traditional approaches was rather small ($d = .15$). Yet other research shows that PBL has particular advantages in comparison with other instructional approaches regarding the acquisition of skills and application-oriented knowledge (Dochy, Segers, van den Bossche, & Gijbels, 2003; Gijbels, Dochy, van den Bossche, & Segers, 2005; Schmidt, van der Molen, Te Winkel, & Wijnen, 2009). Since EBP is supposed to be applied in real-world settings (it is application-oriented knowledge), student-centered approaches like PBL might be more effective in teaching EBP than strongly teacher-centered approaches like DI. Indeed, Tian, Liu, Yang, and Shen (2013) found PBL to be more effective in teaching the process of EBP compared to a lecture-based

approach in a randomized controlled trial with medical postgraduates ($n = 103$).

Knowledge application. Since EBP can be considered application-oriented knowledge, it is important to explore how knowledge is applied within learning processes, for example, working with a fictional case or with real clients (or if knowledge is applied at all). The concept of "situated cognition" tackles the importance of knowledge application during the learning process. The basic idea of situated cognition is not to focus only on isolated aspects like cognition, but take into account the individuals and their actions as well as the situation in which practice takes place (Wilson & Myers, 2000). Proponents of situated cognition such as Lave (1988) assume that during the learning process knowledge cannot be decontextualized, transmitted, and then applied in another context (see Gruber, Law, Mandl, & Renkl, 1996, for an overview of situated learning models). She found that skills learned in informal environments are rarely generalized but remain connected to the contexts and the circumstances in which they are acquired. She emphasized the importance of everyday practice and the necessity to embed learners in social communities that support participation and increasingly independent application of skills in relevant settings (see more recent research concerning the importance of deliberate practice in enhancing expertise such as Rousmaniere, Goodyear, Miller, & Wampold, 2017).

Effects of EIs

Much research is interested in studying the effects of certain EIs on desired outcomes. An effect is the difference between what happened when people received an intervention and what would have happened if they had not received it (Shadish, Cook, & Campbell, 2002). One important outcome is knowledge acquisition that may be declarative and/or procedural (Anderson, 1996). Declarative knowledge (knowing what) refers to knowledge about facts, concepts, and principles. Procedural knowledge (knowing how) refers to skills and actions. Researchers are also interested in effects of EIs on other variables such as learner's motivation to engage with the subject matter (e.g., Ruzafa-Martínez, López-Iborra, Armero Barranco, & Ramos-Morcillo, 2016). The development of standardized instruments to measure social workers' attitudes toward and intentions to use EBP suggest that motivation toward EBP is an important construct related to EIs in social work (Aarons, 2004; Aarons et al., 2010; Rubin & Parrish, 2010). Finally, the learner's satisfaction with an EI is also an outcome that is often measured in EI studies.

Quality of Empirical Intervention Studies

To determine the effectiveness of an EI on relevant dependent variables, it is important to consider the methodological quality of related empirical studies. Study quality can be operationalized at different levels including rigor in design and reliability and validity measures. Both concerns are affected by risk of bias which we also discuss.

Rigor in design. Studies that lack a controlled design can be problematic in identifying effects and do not support strong causal inferences (Shadish et al., 2002). This does not mean that discovery of important aspects of learning is restricted to well-controlled experimental research (Hoadley & van Heneghan, 2012). Yet the inclusion of a control group is a sign of quality with regard to claimed effects, especially for quantitative methods. Nevertheless, Yaffe (2013) notes that most evaluation studies in social work education do not apply a controlled design. Qualitative research usually has other goals than detecting a causal relationship such as reconstructing interpretative patterns or exploring learners' individual adaptations of knowledge. Qualitatively oriented researchers may speculate about what would have happened if a causal factor was missing (Johnson & Christensen, 2013).

Reliability and validity. Reliability refers to how consistently a construct is measured. One way to estimate the reliability of a measure is to examine its internal consistency, how closely items on a measure are related by calculating the Cronbach's α. An alternative is examining stability of a measure by administering this at different times and examining scores. Validity refers to whether a measure actually reflects the construct of interest. Different kinds of validity include content validity (do items accurately reflect the domain of interest?), construct validity including convergent (are two constructs that should theoretically be related in fact related?) and divergent/discriminant validity (are two constructs that should theoretically be not related in fact not related?), criterion validity that includes concurrent validity (relationship between test scores and criterion scores obtained at the same time), and predictive validity (relationship between test scores obtained at one point in time and criterion scores obtained at a later time). Self-report measures may not reflect behavior in real-life settings. Relying solely on learners' perceived learning is problematic since we tend to overestimate our knowledge (Kruger & Dunning, 1999; Snibsøer et al., 2018). Instead, when assessing knowledge and its use, observation of performance in real life or simulated work settings using valid measures is preferable (Johnson & Christensen, 2013). Thus, the "measurement strategy" (performance tests vs. self-report) of a study is a particularly important aspect of validity in our review.

Risk of bias. Bias refers to systemic error in one direction. Factors that may contribute to such bias are, for example, incomplete outcome data (attrition bias) or selective outcome reporting (reporting bias; Higgins & Green, 2011). Risk of bias assessment is closely connected to the type of empirical data used, the theoretical rationale drawn on and the unique circumstances of a study. Different methods to assess risk of bias exist and the method used in a particular review should be selected with reference to the methodological features relevant to the included studies (Liberati et al., 2009).

Objectives and Research Questions

The aim of this study is to describe and review research on EIs used to teach the process of EBP and/or ESTs to social work students and practitioners and their effects on various dependent variables (such as knowledge, motivation, and satisfaction), considering the quality of the studies. We investigated the following research questions:

- **Research Question 1:** What EIs are applied in research on EBP education in social work? Given the potentials of more student-centered approaches regarding the acquisition of application-oriented knowledge, we were particularly interested in the extent to which such approaches have been used in related research. And, was knowledge applied during the learning processes and if so, how?
- **Research Question 2:** What are the effects of these EIs? We are interested in effects regarding the acquisition of both declarative and procedural knowledge, motivation toward EBP, and satisfaction with the EIs.
- **Research Question 3:** What is the methodological quality of the studies? To what extent did studies use controlled designs and reliable and valid measures? We are especially interested in the studies' measurement strategies (performance tests vs. self-report) and their risk of bias.

We carried out a systematic review to answer these questions. Due to the varied means of data collection and analysis in research reports (qualitative, quantitative, and mixed methods) as well as the heterogeneity of designs, samples, and interventions, we did not conduct a meta-analysis. Instead, we provide a narrative synthesis.

Method

We largely followed the *Cochrane Handbook of Systematic Reviews for Interventions* (Higgins & Green, 2011) in conducting our review. However, we used different criteria to estimate risk of bias since the criteria suggested in the handbook focus mainly on well-designed randomized controlled trials that were rare in our sample (see later discussion of risk of bias). We followed the preferred reporting items for systematic reviews and meta-analyses (PRISMA) reporting guidelines regarding pertinent categories (Liberati et al., 2009).

Eligibility Criteria

We included studies that met the following criteria. First, the studies had to be empirical. Second, the studies had to include one or several interventions designed to help participants develop relevant declarative and/or procedural knowledge and/or motivation regarding ESTs and/or the process of EBP (studies that address both approaches are labeled as "Both"). Studies solely interested in learners' satisfaction with a particular EI were not included. Third, the sample used had to consist at least partially of social workers or social work students.

Table 1. Search Terms Used to Find Relevant Studies for a Systematic Review of Educational Intervention Studies to Teach the Process of EBP and/or ESTs in Social Work.

Construct	Search Term(s)
Field	Social Work*
EBP	Evidence N1 based Evidence N1 support Evidence N1 informed 2 OR 3 OR 4
Educational concepts	Teach* Train* Workshop Educat* Curricul* Apprais* Implement* Attitud* Learn* Instruct* Course Foster Facilitat* Appl* 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12 OR 13 OR 14 OR 15 OR 16 OR 17 OR 18 OR 19
Combined Terms	1 AND 5 AND 20

Note. EBP = process of evidence-based practice and empirically supported interventions; N1 = near/*n*, "near" searches for instances of the search term in any order, "*n*" specifies number of terms between the search terms (e.g., "based on evidence" would be a possible result). Syntax (e.g., N/1, near/1) was adjusted due to the requirements of the respective database.

Fourth, only studies in English or German language were included (see Online Appendix Table A1 for a detailed description of the eligibility criteria).

Literature Search

We carried out a bibliographic search to locate relevant articles using search terms grouped into the following categories: *field, EBP,* and *educational concepts* (see Table 1). We used a filter of "peer-reviewed" to identify publications subject to some kind of quality control. We searched in the following databases using the combined terms (see Table 1): *Social Services Abstracts*, *Sociological Abstracts*, *Applied Social Sciences Index and Abstracts*, *SocINDEX*, *PsychINFO*, *ERIC*, *Social Service Citation Index*, *and Social Care Online*. The search was carried out in December 2017. In addition, we handsearched the *Journal of Evidence-Based Social Work* since it is particularly concerned with research on EBP in social work. Furthermore, we handsearched the special issue of 2015 of *Research on Social Work Practice* regarding the Houston Bridging the Research-Practice Gap Symposium, and we also performed a snowball search for relevant articles in the references of already retrieved articles.

Table 2. Interrater Reliability for Eligibility Criteria.

Interrater Reliability/ Screening Rounds	Eligibility Criteria			
	Empirical	Intervention	Sample	EBP
κ	0.78	0.82	0.62 (1.0)[a]	1.0
Screenings	3	8	4	4

Note. EBP = process of evidence-based practice and empirically supported interventions, κ = Cohen's Kappa.
[a]The low κ value for "sample" occurred with only one conflict (interrater agreement was 87.5%) due to binary coding (social work vs. no social work). After achieving consensus for this one case, κ for sample was 1.0.

Study Selection

Two independent coders used the described eligibility criteria to review abstracts of >10% of all potentially relevant articles using a binary code (study to be included vs. study not to be included), until a sufficient interrater reliability (IR; Cohen's Kappa coefficient = κ) was reached. When a screening failed to attain a sufficient IR, conflicting cases were resolved through discussion, the coding scheme was adjusted accordingly and a new screening conducted with a new set of >10% of all articles in an interactive process. Table 2 provides an overview of the eligibility criteria and their respective κ values as well as the screening rounds needed to attain these values.

After a sufficient κ was attained for all eligibility criteria, the remaining articles were coded by the first author. The 52 articles that remained after abstract screening were subject to a full text screening by the first author (see Figure 1 for a visualized description of the inclusion and exclusion process).

Data Extraction

We defined a set of variables (see Table 3) to answer our research questions and extracted respective data from the articles. The procedure of data extraction differed with respect to different variables.

Descriptive variables. Variables that are rather descriptive in nature such as location where a study was conducted or the duration of an EI were not coded but directly extracted.

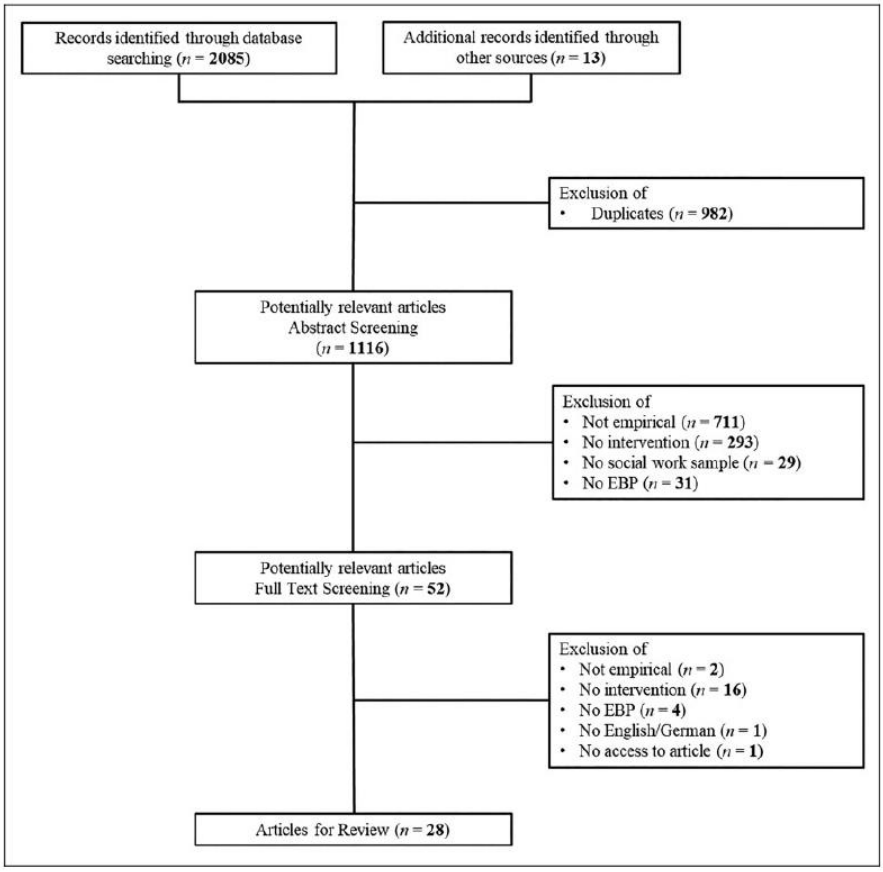


Figure 1. Flow diagram of the inclusion and exclusion process of studies for the review. Two of the 28 articles refer to one study, thus we analyzed 27 studies.

Table 3. Variables.

Variable	Description/Subcodes	IR
Preliminary analysis		
Location	The location where the study was conducted	N/A
Sampling method	Procedure for selecting sampling members	N/A
Sample	The sample must contain at least one social work student (BSW, MSW, PhD) or social worker (e.g., case manager, clinical supervisor, field instructor). It may also contain other professionals (e.g., nurses, psychologists). *Subcodes*: "BSW," "MSW," "PhD," "social worker," "mixed social work" (social work students and social work professionals), "mixed students" (social work students as well as students from other profession/s, e.g., BSW and psychology students), "mixed professionals" (social workers as well as professionals from other profession/s, e.g., social workers and nurses), and "mixed" (students and professionals from social work and other profession/s)	consensus
Sample size	The sample size of a studies' sample	N/A
Gender	*Subcodes*: "exclusively female" (the sample consisted entirely of females), "predominantly female" (75% or more were female), "mixed" (females and males were both below 75%), "predominantly male" (75% or more were male), and "exclusively male" (the sample consisted entirely of males)	$\kappa = 1.00$
Age	The variable "age" refers to the mean age of the sample. If the mean was not provided, the median was extracted. *Subcodes*: "under 30," "between 30 and 40," "between 40 and 50," and "between 50 and 60"	$\kappa = 0.75$
Ethnic diversity	Refers to the question whether a study was based on an ethnically diverse or homogenous sample	N/A
Intervention	Refers to the question whether the educational intervention was a workshop, a lecture, a university course, and so on	N/A
Duration	The duration of a particular educational intervention	N/A
Steps in EBP process	The steps of the EBP process which were addressed by a particular educational intervention	N/A
EST	The EST (e.g., cognitive behavioral therapy, motivational interviewing) addressed by a particular educational intervention	N/A
Research Question 1—EIs		
Instructional approach	*Subcodes*: "teacher" (e.g., a teacher-centered lecture in which an instructor is attempting to transmit knowledge), "student–teacher" (e.g., a workshop in which instructional parts alternate with problem-based small group work), or "student" (e.g., a student-centered educational intervention that offers a lot a freedom for the learners to apply new knowledge, e.g., with the combination of self-dependent work with real clients)	consensus
Knowledge application	Knowledge application refers to the realness of the EI's case-based content. *Subcodes*: "real-world practice" (e.g., the learner applies content from the EI with actual clients), "case-based" (e.g., the learner applies content from the EI in a fictional social work case), or "not case-based" (the EI does not require to apply knowledge with social work cases)	consensus
Research Question 2—Effects		
Tested declarative knowledge	Declarative knowledge (knowing what) of EBP, measured and subsequently assessed by a third person *Subcodes:* "positive effect" (e.g., increase of the mean from pre- to postmeasurement, citations from participants, or conclusions of researchers that imply a positive effect, e.g., "Through the workshop I realized the importance of empirical research for daily practice"), "no effect" (e.g., conclusions of researchers that no differences between pre- and postmeasurement occurred), and "negative effect" (e.g., decrease of the mean from pre- to postmeasurement)	consensus
Tested procedural knowledge	Procedural knowledge (knowing how) of EBP, measured and subsequently assessed by a third person *Subcodes:* See "tested declarative knowledge"	$\kappa = 1.00$
Perceived declarative knowledge	Declarative knowledge (knowing what) of EBP, measured with a self-report instrument (e.g., Questionnaire) *Subcodes:* See "tested declarative knowledge"	$\kappa = 0.62$
Perceived procedural knowledge	Declarative knowledge (knowing what) of EBP, measured with a self-report instrument (e.g., Questionnaire) *Subcodes:* See "Tested Declarative Knowledge"	$\kappa = 0.67$
Motivation	Attitude toward EBP, feasibility, intentions to use, interest in research, ability to make connections between research and practice, and so on *Subcodes:* See "tested declarative knowledge"	consensus

(continued)

Table 3. (continued)

Variable	Description/Subcodes	IR
Satisfaction	Satisfaction, perceived usefulness or helpfulness of the participants with the intervention. *Subcodes:* "positive effect" (e.g., a score above 75% in a postmeasurement, e.g., 8 out of 10 in a 10-point Likert-type questionnaire, citations from participants, or conclusions of researchers that imply satisfaction, e.g., "overall, participants found the training very useful"), "no effect" (e.g., a mediocre score in a postmeasurement, e.g., 5 out of 10 in a 10-point Likert-type questionnaire), and "negative effect" (e.g., a score below 25% in a postmeasurement, e.g., 2 out of 10 in a 10-point Likert-type questionnaire)	consensus
Research Question 3—Quality		
Design	The design was determined by coding the points of measurement (PoM) of the variables of Research Question 2—effects. These variables are depicted in hierarchical order with "tested declarative knowledge" being of the highest and "Satisfaction" of the lowest interest. *Subcodes for the PoM of each variable:* pre, post, follow-up, post follow-up, pre-post, pre follow-up, and pre-post follow-up. These codes are depicted in hierarchical order with "pre" being the least and "pre-post follow-up" being the most desirable subcode. The subcode "prerepeated" was added later on for one particular study. The design was determined through (1) the outcome of highest interest and (2) the most desirable PoM (pre-post over post-only, pre-post follow-up over pre-post)	consensus (motivation) $\kappa = 0.62–1.00$
Measure instrument strategy	Refers to whether the measurement instrument implied a subsequent performance assessment by a third person or not. *Subcodes:* "test" (e.g., observation, knowledge test, vignette), "perception" (e.g., focus groups, interview, Likert-type scale questionnaire) or a combination of "both"	consensus
Analysis paradigm	Refers to the methods of analysis. *Subcodes*: "qualitative" (e.g., content analysis), "quantitative" (e.g., inferential statistics), or "mixed methods" (qualitative and quantitative methods)	$\kappa = 0.82$
Reliability	Measures of reliability (e.g., Cronbach's α) of an applied measurement instrument	N/A
Validity	Refers to the validity of an applied measurement instrument. If a measure was claimed to be valid, we only included information on validity if we could locate other studies in support of these claims	N/A
Risk of bias	The Mixed Methods Appraisal Tool was used to assess the studies' risk of bias (Pluye et al., 2011)	75–100%

Note. EBP = process of evidence-based practice and empirically supported treatments; IR = interrater reliability; κ = Cohen's Kappa; % = interrater agreement; N/A = not applicable (descriptive variable); EI = educational intervention.

Coded variables. To code variables that were not descriptive in nature such as instructional approaches or knowledge application, we developed and iteratively refined a standardized data abstraction form. A number was allocated to each subcode and studies were coded numerically. All studies were coded by two independent coders using >20% of the relevant articles until a sufficient IR ($\kappa > 0.60$) was attained. The remaining articles were coded by the first author. We encountered a great deal of vague or missing descriptions (we contacted eight authors to ask for request additional information and three answered). Thus, all ratings for variables with which we experienced problems to attain a sufficient IR were double coded by two coders based on consensus. Table 3 provides an overview of all coded variables, their operationalization, and their IR.

Risk of bias assessment. The Mixed Methods Appraisal Tool (MMAT) was used to assess the studies' risk of bias (Pluye et al., 2011). The MMAT was developed for use with systematic reviews that include quantitative, qualitative, and mixed method studies. It has been validated with regard to content validity (Pace et al., 2012; Pluye, Gagnon, Griffiths, & Johnson-Lafleur, 2009, 2011). Its reliability ranges from $\kappa = 0.21$ to 1.00. This tool needs improvement, especially regarding qualitative studies (Souto et al., 2015). The MMAT consists of various questions that differ with respect to the data collection and analysis used in a study. It consists of four questions each for qualitative studies, randomized controlled trials, nonrandomized trials, and quantitative (descriptive) studies and three questions for mixed methods studies. Mixed methods studies are rated using these questions as well as qualitative questions and the respective quantitative questions. An exemplary question for mixed method studies would be "Is the integration of qualitative and quantitative data relevant to address the research question (objective)?" Response options are *yes*, no, or don't know (not enough information available). An overall score could be allocated to the studies, ranging from 0 (all questions answered with *no*) to 4 (all questions answered with *yes*). We used binary coding, simply using *yes* or *no*, since *don't know* results in the same rating as *no*. One (>10%) study of each paradigm ($n = 6$; quantitative studies could be quantitative, quasi-experimental or experimental) was rated by two independent coders. Paradigms were determined by the studies

"analysis paradigm" and "design." After two screenings, a sufficient interrater agreement was achieved for each paradigm (range: 75–100%). Remaining studies were coded by the first author. Results based on the MMAT should be treated with caution given the many judgments involved, some based on unclear criteria. Thus, overall scores for the specific studies are not given in this review. However, a summary will be provided to indicate the overall risk of bias of evidence generated with this review.

The variables "sample size," "reliability," and "effects" may have been simplified for evidence aggregation and/or easier readability in the following sense.

Sample size. To determine a study's sample size, we extracted the number of participants who completed the posttest. If a study involved a pretest, we extracted the number of the participants who completed both pre- and posttests. If the study reported more than one outcome of interest, we extracted the smallest of the provided numbers. For example, if a study measured "motivation" of 34 participants and also "satisfaction" of 31 participants, 31 was extracted as the sample size. The same was done for follow-up sample sizes.

Reliability. If multiple values for measurements of internal consistency were reported for subscales relevant to a single dependent variable (DV), the range was reported (e.g., for feasibility, $\alpha = .76$; attitude, $\alpha = .89$; and intentions to use, $\alpha = .63$, we reported $\alpha = .63$–$.89$ for motivation). If multiple internal consistency values were reported for various points of measurement, we computed the mean (e.g., pretest, $\alpha = .90$; posttest, $\alpha = .89$; follow-up, $\alpha = .93$; then we report, $\alpha = .91$).

Effects. With respect to "effects" derived from quantitative results, we extracted the reported means, the standard deviations, and p values (see Online Appendix Table A2). However, note that significance testing is increasingly criticized because of misleading implications (Wasserstein, Schirm, & Lazar, 2019). If studies reported several items (e.g., from a survey) that referred to the same DV, we calculated the mean of the items. If n was reported for each item, we computed a weighted pooled mean. Regarding qualitative results, we extracted quotes from participants or conclusions of researchers that imply an effect on a DV (e.g., "Through the workshop I realized the importance of empirical research for daily practice"). We summarized all extracted data as positive (+), no (o), and negative (−) effect (see Table 5). Regarding the variable satisfaction, (+) was interpreted as a high, (o) as a medium, and (−) as a low level of satisfaction. Some studies had additional objectives that went beyond the examination of the EIs' effects on social workers. For instance, some studies were interested in clients' perceptions of an EST delivered after providers received training using a particular EST. However, only objectives, measures, and effects regarding the facilitation of EBP knowledge, motivation, and/ or satisfaction with the EI were extracted for this review.

Variables

To answer our research questions, the articles included in this analysis were coded with respect to a broad range of variables. Regarding Research Question 1, we coded them with respect to the instructional approach of the respective EI and how (or if) learners had to apply EBP knowledge (knowledge application). With respect to Research Question 2, we coded descriptive as well as procedural EBP knowledge, motivation toward EBP, and the learners' satisfaction with the EI. Concerning Research Question 3, we coded the studies' design, their measure instrument strategy (whether data collected by a measure instrument referred to a performance test or the learners' self-assessment of their knowledge) as well as the studies' analysis paradigm (the methods of analysis). Furthermore, we coded the articles with respect to a number of background variables, to characterize the studies' broader circumstances in a preliminary analysis (see Table 3). All variables are described in more detail in Table 3.

Results

Study Selection

Our search across the different databases revealed 2,085 hits. Handsearching the *Journal of Evidence-Based Social Work* revealed another three potentially relevant articles. Furthermore, hand-searching the special issue of 2015 of *Research on Social Work Practice* regarding the Houston Bridging the Research-Practice Gap Symposium revealed no potentially relevant articles and 10 additional articles were identified with a snowball search. We ended up with 2,098 potentially relevant articles. Figure 1 provides a detailed overview of the inclusion and exclusion process.

One reason for the large discrepancy between the initially identified articles and the final sample lies in the large number of duplicates ($n = 982$), resulting in a search in eight different databases. Another reason might be our broad inclusion criteria yielding many conceptual articles. One article was only available in Spanish. The first author of this article was contacted via e-mail and she confirmed that no English version was available. In the case of one article that was not accessible to us, we contacted all three authors by e-mail. No one replied. Thus, we ultimately analyzed 28 articles that refer to 27 studies (two different articles refer to the same study).

Preliminary Analysis

Fifteen (55.5%) studies were conducted solely with social workers and/or social work students. Eleven studies (40.7%) did not provide any information on the age of the participants and six (22.2%) did not provide information on gender. Table 4 provides an overview of the samples including information on age and gender for studies that represent the two different approaches (the process of EBP and ESTs).

Twenty-three studies were conducted in the United States, three in the UK, and one in Israel (see Table 5). All studies used

Table 4. Sample Characteristics.

Sample Characteristics	Total, *n* (%)	EBP Approaches		
		EBP, n (%)	ESTs, *n* (%)	Both, *n* (%)
All studies	27 (100)	13 (48.1)	12 (44.4)	2 (7.4)
Sample				
Social work				
MSW	6 (22.2)	4 (14.8)	2 (7.4)	0
Social worker	5 (18.5)	4 (14.8)	1 (3.4)	0
BSW	3 (11.1)	1 (3.7)	2 (7.4)	0
Mixed social work	1 (3.7)	0	0	1 (3.7)
Mixed professions				
Mixed professionals	9 (33.3)	4 (14.8)	5 (18.5)	0
Mixed	1 (3.7)	0	1 (3.7)	0
Mixed students	1 (3.7)	0	1 (3.7)	0
Unknown	1 (3.7)	0	0	1 (3.7)
Age				
Under 30	5 (18.5)	2 (7.4)	3 (11.1)	0
Between 30 and 40	5 (24.1)	1 (3.7)	3 (11.1)	1 (3.7)
Between 40 and 50	6 (22.2)	4 (14.8)	2 (7.4)	0
Unknown	11 (40.7)	6 (22.2)	4 (14.8)	1 (3.7)
Gender				
Exclusively female	1 (3.7)	0	1 (3.7)	0
Predominantly female	14 (51.9)	6 (22.2)	7 (25.9)	1 (3.7)
Mixed	6 (22.2)	2 (7.4)	4 (14.8)	0
Unknown	6 (22.2)	5 (18.5)	0	1 (3.7)

Note. EBP = process of evidence-based practice; ESTs = empirically supported treatments.

nonprobability sampling. Sixteen studies used convenience samples, 10 studies used purposive samples (participants fulfilled certain eligibility criteria), and one study did not provide information on its sampling method. Nineteen studies provided information on the ethnicity of participants. All used diverse samples. The majority of the studies evaluated a university course ($n = 10$). Intervention time ranged from 5.5 hr to 9 months for self-regulated trainings and from 7 hr up to one semester for courses and workshops. Five studies focused on enhancing skills in the process of EBP, addressing all five steps in the process. Another five studies, two of which attempted to facilitate both the process of EBP and ESTs, addressed steps one to four in the process. Two studies addressed only Step 3 (critical appraisal) and another two studies addressed Step 4 to some extent (applying research evidence in practice).

Main Characteristics

In line with the PRISMA guidelines (Liberati et al., 2009), Table 5 provides an overview of the study characteristics.

Research Question 1: EIs

Our first research question concerned characteristics of EIs focused on, including instructional approaches and knowledge application. As for instructional approaches, a "student–teacher" approach was used in 15 (55.6%) of the EIs and a "student" approach in 5 (18.5%) meaning that 74.1% of the EIs entailed, at least to some extent, student-centered elements. Two thirds of these studies concerned EBP as a process, focusing on certain steps in the process (see Figure 2).

With respect to knowledge application such as working with a fictional case in classroom, working with simulated or real clients and so on, 13 (48.2%) studies asked learners to apply knowledge in real-world practice. Five (18.5%) studies did not use any case-based application (see Figure 3).

Research Question 2: Effects of EIs

Studies that addressed EBP as a process primarily measured motivation (12 effects, 26.4%) and perceived procedural knowledge (11 effects, 24.2%; see Table 6). Two negative effects were reported, both for motivation and both occurred after a semester-long research course, one with real-world knowledge application (Bender, Altschul, Yoder, Parrish, & Nickels, 2014) and one with case-based knowledge application (Smith, Cohen-Callow, Harnek-Hall, & Hayward, 2007).

For studies that addressed ESTs, 43 (95.6%) of the 45 coded effects were positive. Table 6 provides an overview of the coded effects with respect to the different EBP approaches.

Research Question 3: Study Quality

The third research question concerned quality of studies, specifically their designs, reliability, and validity of measures and risk of bias. Three (11.1%) studies were "qualitative," six (22.2%) "mixed methods" and 18 (66.7%) "quantitative." Twenty-one (77.8%) studies used a one-group design, 4 of

Table 5. Main Characteristics.

Author (Year)	EBP	Main Characteristics (Objective, Location, Sample Size [Follow-Up])	Design	Educational Intervention (Duration)	Instructional Approach, Knowledge Application	Coded Effect: Measure Instrument (Reliability/Validity)
Ager et al. (2005)	ESTs	Foster declarative and procedural motivational enhancement therapy (MET) knowledge, the United States, $n = 175$ (118)	Three-group[a] pre-post follow-up	Workshop (2 days + 4-hr follow-up)	Student–teacher, real world	(+) TDK: MCT ($\alpha = .87$, $r = .61$) (+) TPK: Vignettes (ICC = .68) (+[2]) PPK: LSQ ($\alpha = .65$, $r = .48$–.54) (+) M: LSQ ($\alpha = .82$, $r = .69$)
Bellamy et al. (2013)[b]	EBP	Foster EBP attitudes, understanding, knowledge and skills (article: report participants perceptions of training outcomes, EBP barriers and promoters), the United States, $n = 15$	One-group pre-post	Training in university–agency partnership context (2 days to 3 months)	Student–teacher, case based	(+) M: Field notes, Focus Groups
Bledsoe-Mansori et al. (2013)[b]		Foster EBP attitudes, understanding, knowledge and skills (article: report acceptability of the EI), the United States, $n = 15$				(+) S: Field Notes, Focus Groups
Bender, Altschul, Yoder, Parrish, and Nickels (2014)	EBP	Foster EBP attitudes, familiarity, self-efficacy, current/intended use, and belief that EBP is feasible in the real world, the United States, $n = 152$	Two-group pre-post	Integrating EBP process material in a program evaluation course (1 semester)	Student–teacher, real world	(+) PDK: modified EBPPAS ($\alpha = .89$) (+) PPK: modified EBPPAS ($\alpha = .90$) (−/o/+) M: modified EBPPAS ($\alpha = .67$ to .84) EBPPAS: Content, criterion, and factorial validity
Daniel, Torres, and Reeser (2016)	ESTs	Teach hospice social workers and chaplains an evidence-based music protocol, the United States, $n = 10$	One-group pre-post	Unknown	Unknown, unknown	(+) TDK: MCT ($r = .84$, content validity) (+) PPK: Journal entries (+) M: Focus Group
Dauenhauer, Glose, and Watt (2015)	ESTs	Demonstrate knowledge and skills related to the matter of balance evidence-based falls management program, the United States, $n = 16$	One-group pre-post	University course (one semester)	Student, real world	(+) TDK: MCT (3 subscales, $\alpha =$ 0–.533)
Ducharme, Rober, and Wharff (2015)	ESTs	Develop interns' ability to critically evaluate research and to teach specific manualized interventions (cognitive behavioral therapy [CBT], psychodynamic treatment, interpersonal therapy [IPT]), the United States, $n = 12$	One-group pre-post	University course (one semester)	Student–teacher, case based	(+[3]) PDK: OEQ (+) PPK: LSQ ($\alpha = .92$) (+) S: OEQ
Gromoske and Berger (2017)	EBP	Foster social workers' knowledge of, attitudes toward, views of feasibility, intentions to use, and current use of the EBP process, the United States, $n = 45$ (45)	One-group pre-post follow-up	Workshop (replication of Rubin and Parrish without opinion leader) (7 hr)	Student–teacher, case based	(+) PDK: EBPPAS-short ($\alpha =$ .91[c]) (+) PPK: EBPPAS-short ($\alpha =$ .87[c]) (o/+[2]) M: EBPPAS-short (3 subscales, $\alpha =$ .79[c]–.87[c]) EBPPAS-short: Content, criterion, and factorial validity

(continued)

Table 5. (continued)

Author (Year)	EBP	Main Characteristics (Objective, Location, Sample Size [Follow-Up])	Design	Educational Intervention (Duration)	Instructional Approach, Knowledge Application	Coded Effect: Measure Instrument (Reliability/Validity)
Hagell and Spencer (2004)	EBP	Keeping social care staff informed about research findings, UK, $n = 20$ (20)	One-group post-follow-up	Audio tapes with summarized research knowledge (self-regulated, up to 3 months)	Teacher, not case based	(o) PPK: Focus Group (o) M: Focus Group
Hohman, Pierce, and Barnett (2015)	ESTs	Decrease students' use of dysfunctional communication skills and increase motivational interviewing skills, the United States, $n = 137$	One-group pre-post	University course (one semester)	Student–teacher, case based	(+) TPK: Vignettes ($\alpha = .76^{c}$, ICC = .62–.91) (+) TPK: Vignettes ($\alpha = .87$, ICC = .62–.91)
Holmes (2008)	EBP	Gain a deeper understanding of the relevance and potential application of secondary research within participants' social work practice, as well as becoming familiar and interested in accessing and reading research articles, UK, $n = 6$	One-group post only	Teaching course (3 months)	Student–teacher, not case based	(+) M: OEQ
Kobak, Mundt, and Kennard (2015)	ESTs	Teach cognitive behavior therapy, the United States, $n = 8$	One-group pre-post	Online CBT tutorial (self-regulated, about 5.5 hr)	Teacher, real world	(+) TDK: MCT ($\alpha = .82$) (+) PDK: LSQ (+) PPK: LSQ ($+^{2}$) S: RS ($\alpha = .86$) LSQ: Construct validity
Leathers and Strand (2013)	ESTs	Increase participants' knowledge and attitude about EBPs, the United States, $n = 18$ (18)	rand. Two-group pre-post-follow up	Both groups: Web training (30 min) + access to universities online journals (3 months) EG: Access to practisewise	Student, real world	(+) TDK: MCT ($r = .56$, discriminant validity) (−/+)M: EBPAS ($\alpha = .75$) (+) S: LSQ ($\alpha = .75$)
Leathers, Spielfogel, Blakey, Christian, and Atkins, (2016)	ESTs	Promote the use of an EBP to address child behavior problems, the United States, $n = 57$ (57)	rand. Two-group prerepeated	Both groups: Training of a parent management EBP (16 hr) EG: Change agent (9 months)	Student–teacher, real world	($+^{2}$) PPK: CQ
Lopez, Osterberg, Jensen-Doss, and Rae (2011)	ESTs	Foster attitudes and use of an EBP (behavioral parent training, BPT) in the context of a system level mandate, the United States, $n = 21$ (26)	One-group pre-post follow-up	Workshop (2 days)	Teacher, unknown	($+^{4}$) PPK: LSQ (4 subscales, all $\alpha > .89$) (o) M: EBPAS ($\alpha = .43$–.95) (+) M: LSQ ($\alpha = .62$–.81) (+) S: LSQ ($\alpha = .62$–.81)
Martin, Waites, Hopp, Sobeck, and Agius (2013)	Both	Teach service providers about the benefits of EBHP planning and implementation strategies for older adults, the United States, $n = 18$	One-group pre-post	Training session (2 days)	Student–teacher, not case based	(+) TDK: MCT (+) S: OEQ
Matthieu, Carter, Casner, and Edmond (2016)	EBP	Align practicum instructors' EBP knowledge with content taught in MSW curriculums and enhance abilities to assist students in applying the EBP process in their field placements, the United States, $n = 186$	One-group pre-post	Workshop (1 day)	Student–teacher, case based	(+) PDK: EBPPAS ($+^{2}$) M: EBPPAS EBPPAS: Content, criterion, and factorial validity

(continued)

Table 5. (continued)

Author (Year)	EBP	Main Characteristics (Objective, Location, Sample Size [Follow-Up])	Design	Educational Intervention (Duration)	Instructional Approach, Knowledge Application	Coded Effect: Measure Instrument (Reliability/Validity)
Parrish and Rubin (2011)[a]	EBP	Improve practitioner self-efficacy with and knowledge, attitudes, beliefs regarding the feasibility and intentions and behaviors regarding their adoption and implementation of/ about EBP, the United States, $n = 69$ (61)	One-group pre-post	Workshop with opinion leader (7 hr)	Student–teacher, case based	(+) TDK: MCT (+) PDK: EBPPAS ($\alpha = .91$) (+) PPK: EBPPAS ($\alpha = .86$) (+[3]) M: EBPPAS ($\alpha = .63$–.86) EBPPAS: Content, criterion, and factorial validity
Peterson, Phillips, Bacon, and Machunda (2011)	EBP	EBP, develop students' familiarity with EBP concepts and improve their abilities to search for and use interventions from the professional literature, the United States, $n = 81$	One-group pre-post	BSW course (12 hr)	Student, real world	(+) PDK: LSQ (+[2]) PPK: LSQ (+) M: OEQ
Ronen (2005)	ESTs	To train students in cognitive-behavioral intervention, Israel, $n = 30$	One-group post-only	University course (two semesters) + leading a group intervention in practice (12 weekly sessions + two follow-ups, 75 min each)	Student, real world	(+) PDK: OEQ (+) PPK: OEQ (+) M: OEQ (+) S: LSQ ($\alpha = .82$), OEQ
Sacco et al. (2017)	ESTs	Foster MSW students "screening, brief intervention, and referral to treatment" (SBIRT) knowledge, confidence, skills, and use, the United States, $n = 58$ (33)	One-group pre-post follow-up	Training with standardized clients (15 hr)	Student–teacher, case based	(+) TDK: MCT ($\alpha = .48$[c] knowledge questions and .96[c] case examples) (+) TPK: Observation in role-play (ICC = .82–.96) (+[3]) PPK: LSQ (3 subscales, $\alpha = .95$–.96) (+) PPK: LSQ ($\alpha = .91$[c])
Salcido (2008)	EBP	Teach EBP and enable students to translate theoretical and empirical content into assessment and intervention strategies using management and planning methods, the United States, $n = 38$	One-group pre-post	Workshop (2 days) + macro practice course (MPC) + research course (15 weeks)	Student–teacher, real world	(+) PDK: EBPAS (+) PPK: LSQ (+) S: LSQ
Salloum and Smyth (2013)	ESTs	To assist with implementing a manualized treatment (grief and trauma intervention), the United States, $n = 12$	One-group post-only	Podcasts (self-regulated, 1 day to 3 weeks)	Teacher, real world	(+) S: Interview ($\kappa = .86$)
Smith, Cohen-Callow, Harnek-Hall, and Hayward (2007)	EBP	Foster student attitudes toward and aptitude in using research evidence, the United States, $n = 77$	One-group pre-post	Research methods course (one semester)	Student–teacher, case based	(o) TPK: MCT (+) PPK: LSQ (−) M: LSQ[d]
Straussner et al. (2006)	EBP	Foster the competencies to critically evaluate and apply information from research articles to clinical supervision practice, the United States, $n = 29$	One-group pre-post	Online course (8 weeks)	Teacher, real world	(+) PDK: LSQ (+) PPK: LSQ

(continued)

Table 5. (continued)

Author (Year)	EBP	Main Characteristics (Objective, Location, Sample Size [Follow-Up])	Design	Educational Intervention (Duration)	Instructional Approach, Knowledge Application	Coded Effect: Measure Instrument (Reliability/Validity)
Tennille, Solomon, Brusilovskiy, and Mandell (2016)	Both	To impart EBP skills while enhancing attitudes toward and intentions to practice both the process of EBP and designated EBPs, the United States, $n = 72$ (72)	rand. Two-group pre-post follow-up	Educational project with dyads (student + field instructor; 14 weeks)	Student, real world	(+) PDK: EBPPAS-short ($\alpha = .89$) (+) PPK: EBPPAS-short ($\alpha = .89$) (−/+[2])M: EBPAS ($\alpha = .76$), EBPPAS-short ($\alpha = .89$) EBPPAS-short: Content, criterion, and factorial validity
Webber, Currin, Groves, Hay, and Fernando (2010)	EBP	Delivering research methods and critical appraisal skills training to social workers, UK, $n = 15$	Two-group pre-post	E-learning research methods course (12 weeks)	Teacher, not case-based	(o) TDK: Concept Mapping (+[2]) PPK: LSQ, Focus Group (+) S: Focus Group
Wong (2017)	EBP	To prepare MSW students for appropriate decision-making strategies in working with multicultural client populations, the United States, $n = 80$	One-group post-only	EBP module at university (9 hr)	Student–teacher, real world	(+) TPK: Assignment (+) PDK: OEQ ($\kappa = .87$) (+) PPK: OEQ ($\kappa = .87$) (+) M: OEQ ($\kappa = .87$) (+) S: LSQ

Note. +[n] Signifies number of coded effects. In these cases, effects for more than one construct that refer to one of the DVs (e.g., "attitude" and "intentions to use" subsumed to "motivation") were measured and reported in the respective study. See Table A2 in Online Appendix for a detailed description. EBP = process of evidence-based practice; ESTs = empirically supported treatments; Both = process of evidence-based practice and empirically supported treatments; (+) = positive effect; (o) no effect; (−) negative effect; EBPAS = Evidence-based Practice Assessment Scale; EBPPAS = Evidence-based Practice Process Assessment Scale; TDK = tested declarative knowledge; TPK = tested procedural knowledge; PDK = perceived declarative knowledge; PPK = perceived procedural knowledge; M = motivation; S = satisfaction; MCT = multiple choice test; LSQ = Likert-type Scale Questionnaire; RS = rating scale; OEQ = open-ended questionnaire; CQ = closed questionnaire; α = internal consistency; r = test–retest reliability; ICC = intraclass correlation coefficient; κ = Cohen's Kappa; EG = experimental group; CG = control group.
[a]One group received a delayed intervention (after T2). [b]Articles refer to the same study. [c]Mean value, computed from n values (n = points of measurement with reference to the study design). [d]The study reports a positive effect on "motivation," however, due to our coding scheme, we allocated 7 of 11 items of the respective survey to the construct "perceived procedural knowledge." An example of an effected item would be "evaluating sample size adequacy." Afterward, three of the remaining four items were allocated to the construct "motivation." An example of an affected item would be "original research is confusing." Means were reported for each item and we summarized the three items with a pooled weighted mean.

14

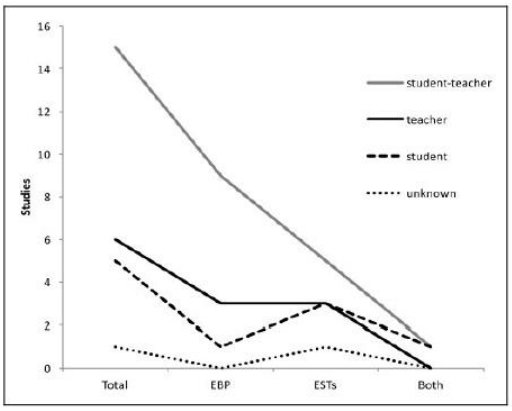


Figure 2. Line graph of instructional approaches.

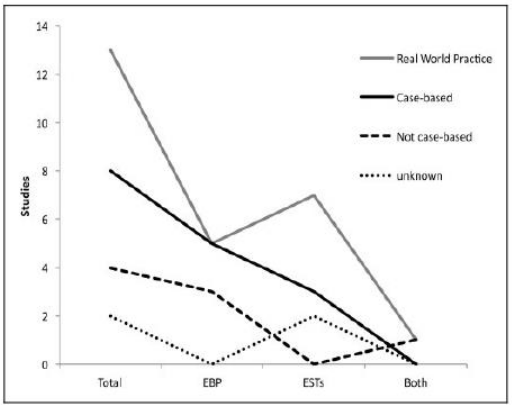


Figure 3. Line graph of knowledge application.

which (two concerned the process of EBP and two EBPs) used only postmeasurements and 8 included follow-ups. About half of the studies applied a one-group pre-post design (48.1%) followed by one-group post-only (14.8%) and one-group pre-post follow-up (11.1%). The designs were evenly distributed among the two EBP approaches. Only six studies (22.2%) used a controlled design. Figure 4 provides an overview of the studies' designs.

Regarding the reliability and validity of measurement instruments, 38 (67.9%) of 56 measurement instruments were quantitative such as use of a Likert-type scale and 18 (32.1%) were qualitative such as an interview. Of the 38 quantitative instruments, 21 (55.3%) provided data regarding internal consistency and 5 (13.2%) provided data concerning test–retest reliability. For eight (21.1%) quantitative instruments, some sort of validity was mentioned. Two (11.1%) of the 18 qualitative instruments provided a value for internal consistency and 6 (33.3%) provided data regarding interrater reliability. Others provided no such information. With regard to measurement, only one (1.8%) measure was a performance test that was based on observation (in role-play; Sacco et al., 2017). Twenty-five (92.6%) studies based their measures solely (51.9%) or partly (40.7%) on self-report data. Figure 5 provides an overview of the studies' measurement strategies, that is, whether the participants' knowledge was actually tested (e.g., multiple choice test) or if they were asked to provide a self-assessment of their knowledge, motivation, and satisfaction (e.g., a Likert-type scale questionnaire).

As for the risk of bias assessment, 1 study scored 0, 4 studies scored 1, 12 studies scored 2, and 12 studies scored 3. No study received an optimal rating of 4. Overall, 15 (55.5%) out of 27 studies scored 0, 1, or 2 (range 0–4).

Table 6. Coded Effects.

	EBP Approach								
	EBP, *n* (%)			ESTs, *n* (%)			Both, *n* (%)		
All Effects	45 (100)			45 (100)			7 (100)		
Effect	+	o	−	+	o	−	+	o	−
TDK	1 (2.2)	1 (2.2)	0	6 (13.2)	0	0	1 (14.3)	0	0
TPK	1 (2.2)	1 (2.2)	0	2 (4.4)	0	0	0	0	0
PDK	8 (17.6)	0	0	5 (11.0)	0	0	1 (14.3)	0	0
PPK	11 (24.2)	1 (2.2)	0	16 (35.2)	0	0	1 (14.3)	0	0
M	12 (26.4)	3 (6.6)	2 (4.4)	7 (15.4)	1 (2.2)	1 (2.2)	2 (28.6)	0	1 (14.3)
S	4 (8.8)	0	0	7 (15.4)	0	0	1 (14.3)	0	0
Total	37 (81.4)	6 (13.2)	2 (4.4)	43 (94.6)	1 (2.2)	1 (2.2)	6 (85.5)	0	1 (14.3)

Note. EBP = process of evidence-based practice; ESTs = empirically supported treatments; TDK = tested declarative knowledge; TPK = tested procedural knowledge; PDK = perceived declarative knowledge; PPK = perceived procedural knowledge; M = motivation; S = satisfaction; + = positive effect; o = no effect; − = negative effect.

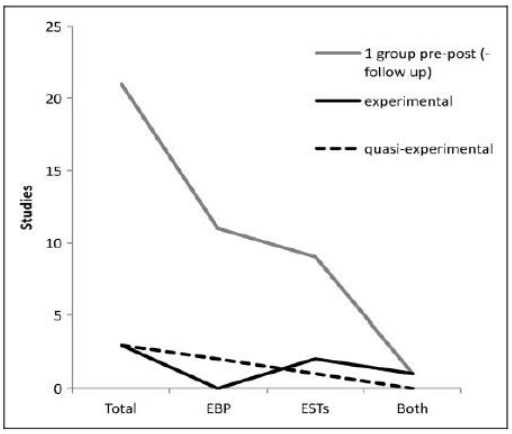


Figure 4. Line graph of study designs. One-group designs were summarized (pre-post, post-only, and with follow-up measurements) to provide a more accessible overview. The same is true for two two-group pre-post and a three-group pre-post follow-up (quasi-experimental) and two randomized two-group pre-post follow-up and a randomized two-group prerepeated (experimental).

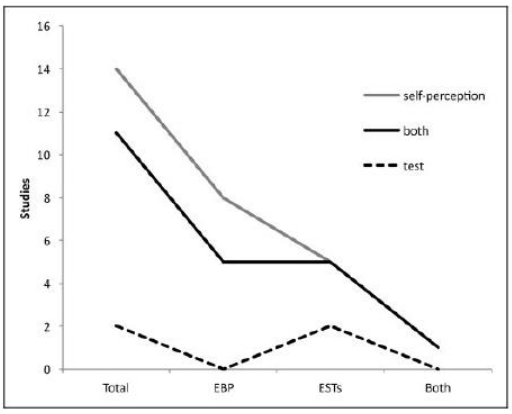


Figure 5. Line graph of the measurement strategies.

Discussion

The aim of this article was to provide a comprehensive overview of empirical studies concerned with supporting social work students and/or professionals in their development and application of EBP. We distinguished between two EBP approaches, namely the process of EBP and ESTs. Our main goals were to find out (1) what kinds of interventions have been used so far to foster EBP in social work, (2) what the effects of these interventions are, and (3) to assess the methodological quality of those studies.

EIs and Their Effects

Research Questions 1 and 2 concerned the conceptualization of EIs and their effects in order to find out how to teach the process of EBP and/or ESTs in social work in an effective way. Studies predominantly applied a guided student-centered instructional approach. This approach is viewed favorably for education in the process of EBP (Straus et al., 2011; Tian, Liu, Yang, & Shen, 2013). Based on a meta-analysis that supports the effectiveness of PBL regarding facilitation of application-oriented knowledge and skills (e.g., Dochy et al., 2003), this focus on student-centered teaching seems warranted. Most studies requested participants to apply EBP knowledge in real-world settings. Learning effects reported as a result of using guided student-centered instructional approaches were mostly positive, especially for studies attempting to foster ESTs (94.6%). Other instructional approaches were also reported to be successful. This may suggest to the uncritical reviewer that any kind of intervention may be effective (Dizon, Grimmer-Somers, & Kumar, 2012). However, reliance on self-report data and variable study quality makes it difficult to determine. Notably, there were no measures of actual use of the process of EBP or ESTs in real-life settings or of the fidelity with which ESTs were implemented with one exception. Sacco et al. (2017) assessed fidelity of an EST used with standardized clients. Clearly, more research that includes the use of relevant declarative and procedural knowledge in real-world settings is needed to discover guidelines for teaching both the process of EBP and ESTs.

Assessment of Study Quality

Our third research question addresses study quality. We approached this question in three ways. First, we looked at the designs that were used in the studies we investigated. Only about one fifth of the studies used a controlled design that allowed for comparison of the effects of different types of instruction. The majority of the studies used a one-group pre-post design, followed by a one-group post-only design. Eleven percent of the studies were qualitative and none of which uses a controlled design. Both controlled designs as well as qualitative research studies are important in educational research and both are underrepresented in our sample. As previously noted, studies without a controlled design do not support causal inferences (Shadish et al., 2002). In summary, to date, studies investigating the effects of EIs on EBP in social work leave unanswered questions regarding the best teaching approach, for example, whether the teaching approach they used is superior to alternative approaches.

Second, we looked at the reliability and validity of measures used. Only about 13% of quantitative measures provided data concerning test–retest reliability, about 20% concerning validity, and only one third of qualitative instruments were checked for reliability. To assess declarative and procedural knowledge, self-reports were much more prevalent than performance measures such as multiple choice tests or observation of performance during practice scenarios or in real-life settings. This is problematic for at least two reasons. First, as mentioned earlier, individuals tend to overestimate their knowledge (Kruger & Dunning, 1999; Snibsøer et al., 2018). And second, because the goal of EIs related

to EBP is (or at least should be) to help learners become more proficient in the use of the process of EBP and/or ESTs *in practice*, asking them for whether they *feel to* be equipped with respect to EBP says little about whether they actually are.

Third, we assessed the risk of bias of the investigated studies. More than half of the studies scored 0-2 (range 0-4). Thus, the very positive results need to be treated with caution. In fact, only one study (Smith et al., 2007) included a "test" as well as a "perception" measure regarding the same dependent variable (procedural knowledge). Even though students reported that they knew more about how to critically analyze research, tests of their knowledge showed no improvement for these skills. This result casts further doubts on relying solely on self-report measures for the assessment of declarative and procedural knowledge, which, as we have seen, seems to be the approach taken in most research on the effects of EIs on EBP in social work. Additional research is needed using reliable, valid performance measures of EBP knowledge and skills.

Recommendations

Given the findings of this review, it is difficult to offer recommendations for teaching the process of EBP and/or ESTs in social work. We should draw on related studies in other areas to inform practices in social work.

Even though most studies in the social work context used EIs based on student-centered teaching approaches, we do not know whether these approaches are actually more effective than other approaches, particularly, more teacher-centered approaches. Perhaps certain kinds of learners (e.g., novices) benefit more from teacher-centered instruction, while others such as more advanced students and practitioners would learn more from student-centered instruction. Evidence from other research areas supports this hypothesis (Kalyuga, 2007). Thus, more research is needed in the social work context to discover what kind of teaching methods under what circumstances are most effective in facilitating the use of EBP by students and practitioners.

Nevertheless, social work educators of course cannot wait for this research to be carried out. In planning courses or other kinds of interventions, we therefore recommend them to carefully review and critically appraise the research evidence they want to base their teaching on and also to consider research from other areas. Based on the review of Aglen (2016), it might be valuable to include aspects of critical thinking (e.g., Gambrill, 2013; Gambrill & Gibbs, 2017) in EBP education. Multifaceted approaches (those using a combination of methods like lectures, computer sessions, small group discussions, journal clubs, and assignments) might be more promising than interventions that offer only one of these methods or no intervention (Kyriakoulis et al., 2016; Patelarou et al., 2017).

Limitations and Conclusions

First, even though all coding was based on reliability checks through double coding, coding still remains a subjective endeavor. Coding was based on published descriptions and some reports failed to provide detailed information, for example, regarding the EI and sample. This might have contributed to moderate interrater reliability values for several variables. Another consequence of lack of detail was that it was not possible to carry out a more specific investigation of the EIs. Second, in order to include qualitative, quantitative, and mixed methods studies, we applied broad operationalizations for the effect variables. This might have contributed to subjectivity in ratings, especially for qualitative results. Also, the classification of effects in "positive," "no," and "negative" is coarse-grained. Third, the broad inclusion criteria used in our review resulted in a study sample including a wide variety of EIs and designs making comparison a challenge. Fourth, the almost exclusively positive results reported make it difficult to discover the most effective training methods for EBP in social work. Fifth, more than half of the studies achieved low scores (0, 1, or 2 out of 4) on risk of bias. Sixth, most studies relied on self-reports. Thus, results of research on EBP education in social work need to be treated with caution. We need more studies using controlled designs with measures that focus on performance rather than self-report.

To conclude, much remains to be done to make informed decisions regarding the design of EIs and measurement of their effects. We hope that our study stimulates additional related empirical research.

Acknowledgments

The authors are very grateful for the support of Carolina Stadler in terms of coding training and reliability testing.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research presented in this work was supported by the Elite Network of Bavaria (Project number: K-GS-2012-209).

ORCID iD

Florian Spensberger https://orcid.org/0000-0002-6783-7329

Supplemental Material

Supplemental material for this article is available online.

References

References marked with an asterisk indicate studies included in the systematic review.

Aarons, G. A. (2004). Mental health provider attitudes toward adoption of evidence-based practice: The Evidence-Based Practice Attitude Scale (EBPAS). *Mental Health Services Research*, *6*, 61–74.

Aarons, G. A., Glisson, C., Hoagwood, K., Kelleher, K., Landsverk, J., & Cafri, G. (2010). Psychometric properties and U.S. national

norms of the Evidence-Based Practice Attitude Scale (EBPAS). *Psychological Assessment*, *22*, 356–365. doi:10.1037/a0019188

*Ager, R., Roahen-Harrison, S., Toriello, M. P. H., Morse, P., Morse, E., Carney, L, . . . Kissinger, P. (2005). A study on the effectiveness of a brief motivational enhancement therapy training. In E. Edmundson & D. McCarty (Eds.), *Implementing evidence-based practices for treatment of alcohol and drug disorders* (pp. 15–32). Binghamton, NY: Haworth Medical Press. doi:10.1300/J069v24S01_02

Aglen, B. (2016). Pedagogical strategies to teach bachelor students evidence-based practice: A systematic review. *Nurse Education Today*, *36*, 255–263. doi:10.1016/j.nedt.2015.08.025

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist*, *51*, 355–365. doi:10.1037/0003-066X.51.4.355

Barnett, S., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, *128*, 612–637. doi:10.1037/0033-2909.128.4.612

Barrows, H. S., & Tamblyn, R. M. (1980). *Problem-based learning: An approach to medical education. Springer series on medical education*: v. *1*. New York, NY: Springer.

Bellamy, J. L., Bledsoe, S. E., & Traube, D. E. (2006). The current state of evidence-based practice in social work: A review of the literature and qualitative analysis of expert interviews. *Journal of Evidence-Informed Social Work*, *3*, 23–48. doi:10.1300/J394v03n01_02

*Bellamy, J. L., Fang, L., Bledsoe, S. E., Manuel, J. I., Dinata, E., & Mullen, E. J. (2013). Agency-university partnership for evidence-based practice in social work. *Journal of Evidence-Informed Social Work*, *10*, 73–90. doi:10.1080/15433714.2011.581545

*Bender, K., Altschul, I., Yoder, J., Parrish, D., & Nickels, S. J. (2014). Training social work graduate students in the evidence-based practice process. *Research on Social Work Practice*, *24*, 339–348. doi:10.1177/1049731513506614

*Bledsoe-Mansori, S. E., Manuel, J. I., Bellamy, J. L., Fang, L., Dinata, E., & Mullen, E. J. (2013). Implementing evidence-based practice: Practitioner assessment of an agency-based training program. *Journal of Evidence-Informed Social Work*, *10*, 73–90. doi:10.1080/15433714.2011.581545

*Daniel, R. A., Torres, D., & Reeser, C. (2016). Where words fail, music speaks: A mixed method study of an evidence-based music protocol. *Journal of Evidence-Informed Social Work*, *13*, 535–551. doi:10.1080/23761407.2015.1111825

*Dauenhauer, J. A., Glose, S., & Watt, C. (2015). Design, delivery, and outcomes from an interprofessional fall prevention course. *Gerontology & Geriatrics Education*, *36*, 278–301. doi:10.1080/02701960.2015.1031891

Dizon, J. M. R., Grimmer-Somers, K. A., & Kumar, S. (2012). Current evidence on evidence-based practice training in allied health: A systematic review of the literature. *International Journal of Evidence-Based Healthcare*, *10*, 347–360. doi:10.1111/j.1744-1609.2012.00295.x

Dochy, F., Segers, M., van den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: A meta-analysis. *Learning and Instruction*, *13*, 533–568. doi:10.1016/S0959-4752(02)00025-7

*Ducharme, P., Rober, A., & Wharff, E. (2015). Building confidence in social work interns through an evidence-based practice seminar during field education. *Field Educator*, *5.1*.

Emparanza, J. I., Cabello, J. B., & Burls, A. J. E. (2015). Does evidence-based practice improve patient outcomes? An analysis of a natural experiment in a Spanish hospital. *Journal of Evaluation in Clinical Practice*, *21*, 1059–1065. doi:10.1111/jep.12460

Fischer, F., Hmelo-Silver, C. E., Goldman, S. R., & Reimann, P. (Eds.). (2018). *International handbook of the learning sciences*. New York, NY: Routledge.

Gambrill, E. (1999). Evidence-based practice: An alternative to authority-based practice. *Families in Society: The Journal of Contemporary Social Services*, *80*, 341–350. doi:10.1606/1044-3894.1214

Gambrill, E. (2006). Evidence-based practice and policy: Choices ahead. *Research on Social Work Practice*, *16*, 338–357. doi:10.1177/1049731505284205

Gambrill, E. (2013). *Social work practice: A critical thinker's guide* (3rd ed.). New York, NY: Oxford University Press.

Gambrill, E. (2019). *Critical thinking and the process of evidence-based practice*. New York, NY: Oxford University Press.

Gambrill, E., & Gibbs, L. (2017). *Critical thinking for helping professionals: A skills-based workbook*. New York, NY: Oxford University Press.

Gibbs, L., & Gambrill, E. (2002). Evidence based practice: Counterarguments to objections. *Research on Social Work Practice*, *12*, 452–476. doi:10.1177/1049731502012003007

Gijbels, D., Dochy, F., van den Bossche, P., & Segers, M. (2005). Effects of problem-based learning: A meta-analysis from the angle of assessment. *Review of Educational Research*, *75*, 27–61. doi:10.3102/00346543075001027

Gray, M., Joy, E., Plath, D., & Webb, S. A. (2012). Implementing evidence-based practice. *Research on Social Work Practice*, *23*, 157–166. doi:10.1177/1049731512467072

*Gromoske, A. N., & Berger, L. K. (2017). Replication of a continuing education workshop in the evidence-based practice process. *Research on Social Work Practice*, *27*, 676–682. doi:10.1177/1049731515597477

Gruber, H., Law, L., Mandl, H., & Renkl, A. (1996). Situated learning and transfer. In P. Reimann & H. Spada (Eds.), *Journals of related interest. Learning in humans and machines: Towards an interdisciplinary learning science* (1st ed., pp. 138–188). Oxford, England: Pergamon Press.

Guyatt, G. H. (1991). Evidence-based medicine. *ACP Journal Club*, *114*, A16–A16. doi:10.7326/ACPJC-1991-114-2-A16

*Hagell, A., & Spencer, L. (2004). An evaluation of an innovative audiotape method for keeping social care staff up to date with the latest research findings. *Child and Family Social Work*, *9*, 187–196. doi:10.1111/j.1365-2206.2004.00313.x

Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement* (Reprint ed.). London, England: Routledge.

Haynes, R. B., Devereaux, P. J., & Guyatt, G. H. (2002). Clinical expertise in the era of evidence-based medicine and patient choice. *ACP Journal Club*, *136*, A11–A14.

Higgins, J. P. T. & Green, S. (Eds.). (2011). *Cochrane handbook of systematic reviews of interventions* (Version 5.1.0). Retrieved from http://handbook.cochrane.org

Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review*, *16*, 235–266. doi:10.1023/B:EDPR.0000034022.16470.f3

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, *42*, 99–107. doi:10.1080/00461520701263368

Hoadley, C., & van Heneghan, J. P. (2012). The learning sciences: Where they came from and what it means for instructional designers. In R. A. Reiser (Ed.), *Trends and issues in instructional design and technology* (3rd ed., pp. 53–63). Boston, MA: Pearson.

*Hohman, M., Pierce, P., & Barnett, E. (2015). Motivational interviewing: An evidence-based practice for improving student practice skills. *Journal of Social Work Education*, *51*, 287–297. doi:10.1080/10437797.2015.1012925

*Holmes, C. M. (2008). How can curriculum action research be used to promote post qualifying practice teaching award candidates' interest in research articles? *Social Work Education*, *27*, 695–709. doi:10.1080/02615470701538215

Ioannidis, J. P. A. (2016). The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. *The Milbank Quarterly*, *94*, 485–514. doi:10.1111/1468-0009.12210

Johnson, B., & Christensen, L. (2013). *Educational research: Quantitative, qualitative, and mixed approaches* (5th ed.). Thousand Oaks, CA: Sage.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19*, 509–539. doi:10.1007/s10648-007-9054-3

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*, 75–86. doi:10.1207/s15326985ep4102_1

*Kobak, K. A., Mundt, J. C., & Kennard, B. (2015). Integrating technology into cognitive behavior therapy for adolescent depression: A pilot study. *Annals of General Psychiatry*, *14*, 37. doi:10.1186/s12991-015-0077-8/

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*, 1121–1134.

Kyriakoulis, K., Patelarou, A., Laliotis, A., Wan, A. C., Matalliotakis, M., Tsiou, C., & Patelarou, E. (2016). Educational strategies for teaching evidence-based practice to undergraduate health students: Systematic review. *Journal of Educational Evaluation for Health Professions*, *13*, 34. doi:10.3352/jeehp.2016.13.34

Lave, J. (1988). *Cognition in practice: Mind, mathematics and culture in everyday life*. New York, NY: Cambridge University Press.

*Leathers, S. J., & Strand, T. C. (2013). Increasing access to evidence-based practices and knowledge and attitudes. *Research on Social Work Practice*, *23*, 669–679. doi:10.1177/1049731513491635

*Leathers, S. J., Spielfogel, J. E., Blakey, J., Christian, E., & Atkins, M. S. (2016). The effect of a change agent on use of evidence-based mental health practices. *Administration and Policy in Mental Health*, *43*, 768–782. doi:10.1007/s10488-015-0694-1

Liberati, A., Altman, D., Tetzlaff, J., Mulrow, C., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *PLoS Medicine*, *6*. doi:10.1136/bmj.b2700

*Lopez, M. A., Osterberg, L. D., Jensen-Doss, A., & Rae, W. A. (2011). Effects of workshop training for providers under mandated use of an evidence-based practice. *Administration and Policy in Mental Health*, *38*, 301–312. doi:10.1007/s10488-010-0326-8

Loyens, S. M. M., & Rikers, R. M. J. P. (2011). Instruction based on inquiry. In R. E. Mayer & P. A. Alexander (Eds.), *Educational psychology handbook series. Handbook of research on learning and instruction* (pp. 361–382). New York, NY: Routledge.

*Martin, F., Waites, C., Hopp, F. P., Sobeck, J., & Agius, E. (2013). Enhancing aging services through evidence-based health promotion: A training for service providers. *Journal of Evidence-Informed Social Work*, *10*, 482–493. doi:10.1080/15433714.2012.760296

*Matthieu, M. M., Carter, L., Casner, R. W., & Edmond, T. E. (2016). Training outcomes of field instructors in the evidence-based practice process model. *Journal of Teaching in Social Work*, *36*, 477–489. doi:10.1080/08841233.2016.1242524

Morago, P. (2010). Dissemination and implementation of evidence-based practice in the social services: A UK survey. *Journal of Evidence-Informed Social Work*, *7*, 452–465. doi:10.1080/15433714.2010.494973

Murphy, A., & McDonald, J. (2004). Power, status and marginalisation: Rural social workers and evidence-based practice in multidisciplinary teams. *Australian Social Work*, *57*, 127–136. doi:10.1111/j.1447-0748.2004.00127.x

Pace, R., Pluye, P., Bartlett, G., Macaulay, A. C., Salsberg, J., Jagosh, J., & Seller, R. (2012). Testing the reliability and efficiency of the pilot Mixed Methods Appraisal Tool (MMAT) for systematic mixed studies review. *International Journal of Nursing Studies*, *49*, 47–53. doi:10.1016/j.ijnurstu.2011.07.002

*Parrish, D. E., & Rubin, A. (2011). An effective model for continuing education training in evidence-based practice. *Research on Social Work Practice*, *21*, 77–87. doi:10.1177/1049731509359187

Parrish, D. E., & Rubin, A. (2012). Social workers' orientations toward the evidence-based practice process: A comparison with psychologists and licensed marriage and family therapists. *Social Work*, *57*, 201–210. doi:10.1093/sw/sws016

Patelarou, A., Kyriakoulis, K. G., Stamou, A. A., Laliotis, A., Sifaki-Pistolla, D., Matalliotakis, M., ... Patelarou, E. (2017). Approaches to teach evidence-based practice among health professionals: An overview of the existing evidence. *Advances in Medical Education and Practice*, *8*, 455–464. doi:10.2147/AMEP.S134475

*Peterson, S. M., Phillips, A., Bacon, S. I., & Machunda, Z. (2011). Teaching evidence-based practice at the BSW level: An effective capstone project. *Journal of Social Work Education*, *47*, 509–524. doi:10.5175/JSWE.2011.200900129

Pluye, P., Gagnon, M. P., Griffiths, F., & Johnson-Lafleur, J. (2009). A scoring system for appraising mixed methods research, and

concomitantly appraising qualitative, quantitative and mixed methods primary studies in mixed studies reviews. *International Journal of Nursing Studies*, *46*, 529–546. doi:10.1016/j.ijnurstu.2009.01.009

Pluye, P., Robert, E., Cargo, M., Bartlett, G., O'Cathain, A., Griffiths, F., . . . Rousseau, M. C. (2011). *Proposal: A mixed methods appraisal tool for systematic mixed studies reviews*. Retrieved from http://mixedmethodsappraisaltoolpublic.pbworks.com

Pope, N. D., Rollins, L., Chaumba, J., & Risler, E. (2011). Evidence-based practice knowledge and utilization among social workers. *Journal of Evidence-Informed Social Work*, *8*, 349–368. doi:10.1080/15433710903269149

Renkl, A., Mandl, H., & Gruber, H. (1996). Inert knowledge: Analyses and remedies. *Educational Psychologist*, *31*, 115–121. doi:10.1207/s15326985ep3102_3

*Ronen, T. (2005). Students' evidence-based practice intervention for children with oppositional defiant disorder. *Research on Social Work Practice*, *15*, 165–179. doi:10.1177/1049731504271604

Rousmaniere, T., Goodyear, R. K., Miller, S. D., & Wampold, B. E. (Eds.). (2017). *The cycle of excellence: Using deliberate practice to improve supervision and training*. Hoboken, NJ: John Wiley.

Rubin, A., & Parrish, D. (2007). Views of evidence-based practice among faculty in master of social work programs: A national survey. *Research on Social Work Practice*, *17*, 110–122. doi:10.1177/1049731506293059

Rubin, A., & Parrish, D. (2010). Development and validation of the evidence-based practice process assessment scale: Preliminary findings. *Research on Social Work Practice*, *20*, 629–640. doi:10.1177/1049731508329420

Ruzafa-Martínez, M., López-Iborra, L., Armero Barranco, D., & Ramos-Morcillo, A. J. (2016). Effectiveness of an evidence-based practice (EBP) course on the EBP competence of undergraduate nursing students: A quasi-experimental study. *Nurse Education Today*, *38*, 82–87. doi:10.1016/j.nedt.2015.12.012

*Sacco, P., Ting, L., Crouch, T. B., Emery, L., Moreland, M., Bright, C., . . . DiClemente, C. (2017). SBIRT training in social work education: Evaluating change using standardized patient simulation. *Journal of Social Work Practice in the Addictions*, *17*, 150–168. doi:10.1080/1533256X.2017.1302886

Sackett, D. L., Richardson, W. S., Rosenberg, W., & Haynes, R. B. (1997). *Evidence-based medicine: How to practice and teach EBM*. New York, NY: Churchill Livingstone.

Sackett, D. L., Rosenberg, W. M., Gray, J., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *BMJ: British Medical Journal*, *312*, 71–72.

*Salcido, R. M. (2008). Incorporating evidence-based practice into the macro practice curriculum. *Journal of Evidence-Informed Social Work*, *5*, 623–645. doi:10.1080/15433710802084359

*Salloum, A., & Smyth, K. (2013). Clinicians' experiences of a podcast series on implementing a manualized treatment. *Journal of Technology in Human Services*, *31*, 71–83. doi:10.1080/15228835.2012.738382

Sawyer, R. K. (Ed.). (2009). *The Cambridge handbook of the learning sciences* (Reprint ed.). Cambridge, England: Cambridge University Press.

Schmidt, H. G., van der Molen, H. T., Te Winkel, W. W. R., & Wijnen, W. H. F. W. (2009). Constructivist, problem-based learning does work: A meta-analysis of curricular comparisons involving a single medical school. *Educational Psychologist*, *44*, 227–249. doi:10.1080/00461520903213592

Scurlock-Evans, L., & Upton, D. (2015). The role and nature of evidence: A systematic review of social workers' evidence-based practice orientation, attitudes, and implementation. *Journal of Evidence-Informed Social Work*, *12*, 369–399. doi:10.1080/15433714.2013.853014

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

Slavin, R. E. (2018). *Educational psychology: Theory and practice* (12th ed.). New York, NY: Pearson.

*Smith, C. A., Cohen-Callow, A., Harnek-Hall, D. M., & Hayward, R. A. (2007). Impact of a foundation-level MSW research course on students' critical appraisal skills. *Journal of Social Work Education*, *43*, 481–495.

Snibsøer, A. K., Ciliska, D., Yost, J., Graverholt, B., Nortvedt, M. W., Riise, T., & Espehaug, B. (2018). Self-reported and objectively assessed knowledge of evidence-based practice terminology among healthcare students: A cross-sectional study. *PLoS One*, *13*, e0200313. doi:10.1371/journal.pone.0200313

Souto, R. Q., Khanassov, V., Hong, Q. N., Bush, P. L., Vedel, I., & Pluye, P. (2015). Systematic mixed studies reviews: Updating results on the reliability and efficiency of the Mixed Methods Appraisal Tool. *International Journal of Nursing Studies*, *52*, 500–501. doi:10.1016/j.ijnurstu.2014.08.010

Straus, S. E., Richardson, W. S., Glasziou, P., & Haynes, R. B. (2011). *Evidence-based medicine: How to practice and teach it* (4th ed., Reprint.). Edinburgh, Scotland: Churchill Livingstone Elsevier.

*Straussner, S. L. A., Naegle, M. A., Gillespie, C., Wolkstein, E., Donath, R., & Azmitia, E. C. (2006). The SATOL project: An interdisciplinary model of technology transfer for research-to-practice in clinical supervision for addiction treatment. *Journal of Evidence-Informed Social Work*, *3*, 39–54. doi:10.1300/J394v03n03_04

Sur, R. L., & Dahm, P. (2011). History of evidence-based medicine. *Indian Journal of Urology: IJU: Journal of the Urological Society of India*, *27*, 487–489. doi:10.4103/0970-1591.91438

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory (1st ed.). Explorations in the Learning Sciences, Instructional Systems and Performance Technologies: Vol. 1*. New York, NY: Springer.

Sweller, J., Kirschner, P. A., & Clark, R. E. (2007). Why minimally guided teaching techniques do not work: A reply to commentaries. *Educational Psychologist*, *42*, 115–121. doi:10.1080/00461520701263426

Task Force. (1993). *Task force on the promotion and dissemination of psychological procedures*. Retrieved from http://www.div12.org/sites/default/files/InitialReportOfTheChamblessTaskForce.pdf

Teater, B., & Chonody, J. M. (2018). Identifying as an evidence-based social worker: The influence of attitudes, access, confidence, and education. *Social Work Education*, *37*, 442–457. doi:10.1080/02615479.2017.1421161

*Tennille, J., Solomon, P., Brusilovskiy, E., & Mandell, D. (2016). Field instructors extending EBP learning in dyads (FIELD):

Results of a pilot randomized controlled trial. *Journal of the Society for Social Work and Research*, *7*, 1–22. doi:10.1086/684020

Thyer, B. A., & Myers, L. (2011). The quest for evidence-based practice: A view from the United States. *Journal of Social Work*, *11*, 8–25. doi:10.1177/1468017310381812

Tian, J. H., Liu, A. P., Yang, K. H., & Shen, X. (2013). The effectiveness of problem-based learning on evidence based medicine: A double-blind randomized trial. *Indian Journal of Pharmaceutical Education and Research*, *47*, 14–18.

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond "$p < 0.05$." *The American Statistician*, *73*, 1–19. doi:10.1080/00031305.2019.1583913

*Webber, M., Currin, L., Groves, N., Hay, D., & Fernando, N. (2010). Social workers can e-learn: Evaluation of a pilot post-qualifying e-learning course in research methods and critical appraisal skills for social workers. *Social Work Education*, *29*, 48–66. doi:10.1080/02615470902838745

Wilson, B. G., & Myers, K. (2000). Situated cognition in theoretical and practical context. In D. H. Jonassen & S. M. Land (Eds.), *Theoretical foundations of learning environments* (pp. 57–88). Mahwah, NJ: Erlbaum.

*Wong, R. (2017). Evaluating a teaching module on ethically responsible evidence-based practice decision making in an advanced micro practice course. *Journal of Social Work Education*, *53*, 240–259. doi:10.1080/10437797.2016.1260505

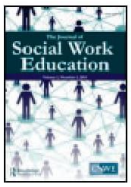
Yaffe, J. (2013). Guest editorial—Where's the evidence for social work education? *Journal of Social Work Education*, *49*, 525–527. doi:10.1080/10437797.2013.820582

7. Study 2 – Effects of Worked Examples and Fallacy Recognition Skills: A Randomized Controlled Trial[8]

This article has been submitted to the Journal of Social Work Education by Taylor & Francis.

[8] This article was co-authored by Ingo Kollar and Sabine Pankofer.

Journal of Social Work Education



Effects of Worked Examples and External Scripts on Fallacy Recognition Skills: A Randomized Controlled Trial

Journal:	*Journal of Social Work Education*
Manuscript ID	USWE-2019-0185.R1
Manuscript Type:	General Manuscript
Content Sections:	Social work practice < Social Work Curricular Area, Educational outcomes evaluation (e.g., course, teacher, program evaluation) < Social Work Curricular - Educational Issues, Evidence-based practice, empirically based practice, practice guidelines, best practices < Other Descriptors

URL: http://mc.manuscriptcentral.com/jswe Email: jswe@cswe.org

Abstract

To provide clients with the best possible treatment, social workers require fallacy recognition skills. Ideally, these skills are already acquired during university education. We conducted an experimental study with a 2x2 factorial design (and an additional baseline condition) with worked examples (with/without) and external scripts (with/without) as scaffolds to support social work students' (N = 130) fallacy recognition skills during case-based reasoning on the exchange of fallacious arguments of social workers in the context of a web-based learning environment. Students significantly improved their fallacy recognition skills over time. During treatment, worked examples and external scripts yielded positive effects when presented in isolation, but not when combined. However, the effects of the two scaffolds could not be maintained in the post-test. Our results show that social work students' fallacy recognition skills in social work decision-making contexts can be facilitated through web-based learning environments on the basis of case-based reasoning. Further systematic research is needed to identify principles for an effective design of additional instructional guidance.

Keywords: Evidence-based Argumentation, Fallacy, Social Work, Social Work Students, Worked Examples, External Scripts, Critical Thinking

Effects of Worked Examples and External Scripts on Social Work Students' Fallacy Recognition Skills: A Randomized Controlled Trial

Problem Statement

Social work clients have a right to receive the best possible treatment for their problem at hand in order to achieve outcomes they value. It has been argued that the process of evidence-based practice (EBP) might contribute to sound decisions regarding the implementation of the best possible treatment as well as to improved outcomes valued by social work clients (Gambrill, 2019). However, empirical evidence suggests that social work practitioners barely engage in EBP (Parrish & Rubin, 2012; Pope, Rollins, Chaumba, & Risler, 2011, van der Zwet, Kolmer, & Schalk, 2016). The ability to engage in critical thinking (CT) is seen as an important prerequisite for EBP (Aglen, 2016; Gambrill, 2019). It is thus reasonable to assume that the promotion of CT contributes to practitioners' engagement in the process of EBP. In a narrow view, CT refers to identifying fallacies (e.g., a mistake in an argument) and bias (Gambrill & Gibbs, 2017). Even though research on social work students' proficiency with respect to fallacy identification appears to be limited, research from other domains has shown that students frequently have severe difficulties in identifying them (e.g., Neuman, 2003; Neuman & Weizman, 2003; Ramasamy & Fung, 2015; Tarnoff, 2010; Weinstock, Neuman, & Tabak, 2004). The concept of fallacies as well as our proposed model of argument will be explained in the following.

Arguments and Fallacies

Deciding which treatment works best for a social work client's problem at hand is not usually an endeavor undertaken by a single social worker. It might rather be based on the exchange of arguments between the client's social worker, colleagues, the client and members of the client's social system. These arguments are conclusions supported by one or more claims (Damer, 2011). Consider the following example:

"It was only three months after Harold got married that he started smoking pot. His wife must have gotten him started on the stuff."

This argument could be broken down into standard form.

"Since Harold recently got married, (claim)

and having a new wife preceded the event of starting to smoke pot, (claim)

Therefore, Harold's wife is the cause of his pot smoking." (conclusion)

Such an argument might be part of a decision-making process regarding the question, "What treatment would be best for this clients' problem?" Decisions for the implementation of a certain treatment depend to a great extent on the criteria that social workers use to evaluate the accuracy of theories and claims of effectiveness (Gambrill, 2013). Thus, if social workers fail to evaluate claims accurately, they might suggest suboptimal or even harmful treatments to their clients (Gambrill, 2011). The fallacy in our example is called "*post hoc ergo propter hoc,*" meaning that whatever precedes another event is the cause of that event. This assumption is implied in the argument and it bridges the gap between the stated claims and the conclusion. However, this assumption is, of course, not always true, hence potentially fallacious. If this fallacy remains undetected, a social worker might recommend Harold to break up with his wife (in order to quit smoking pot). Arguments are prone to numerous fallacies (Damer, 2011).

To avoid the potentially harmful consequences of undetected fallacies, social workers need the skills to identify such fallacies. Damer (2011) claims that a fallacy occurs when an argument or its claims (depending on the criteria) violate one or more of the following five criteria for good argumentation: (1) structure, (2) relevance, (3) acceptance, (4) sufficiency, and (5) rebuttal. The five criteria are briefly explained now with reference to Damer (2011).

Structure. An argument violates the criterion of structure if it uses reasons that contradict each other or the conclusion, or reasons that assume the truth of a conclusion, or if

the argument draws invalid deductive conclusions. An example is, "*The college is very paternalistic in its structures, because it treats students like children."* Note that the premise is the same as the conclusion.

Relevance. A claim violates the criterion of relevance if it is irrelevant for the conclusion or if it appeals to an irrelevant authority or emotional factors. An example is: "*Since I believe in the sanctity of marriage, therefore, I believe that marriage should be between a man and a woman."* Note that the notion of marriage being holy or sacred has no clear connection to the conclusion that it should be between a man and a woman.

Acceptance. A claim violates the criterion of acceptability if it entails either linguistic confusion or an unwarranted assumption. An example is: "*Gambling should be legalized because it is something we can't avoid. It is an integral part of human experience; people gamble every time they get in their cars or decide to get married."* Note that the meaning of the word gambling changes through the course of argumentation. First, it refers to games of chance, then to risks of life.

Sufficiency. A claim violates the criterion of sufficiency if it does not provide reasons that are sufficient in number and weight to justify the acceptance of the conclusion. An example is: *"Vitamin C really works. Every member of my family used to catch at least one cold per year. Last fall, each of us started taking one thousand milligrams of vitamin C a day, and there hasn't been even a sniffle at our house in more than nine months."* Note the (implicit and misleading) claim that what is true for one family is also true for all families.

Rebuttal. An argument violates the rebuttal criterion if it fails to either (1) take relevant counterevidence into account, (2) attacks one's opponent in a personal way, or (3) tries to divert attention away from the actual intention of the argument. An example is: *"No wonder you think sexual promiscuity is all right. You know you've never had a good relationship with a woman. So, it's not strange that you'd resort to recreational sex."* Note

that there is no attempt to rebut a prior argument. Instead, the individual presenting the prior argument is attacked personally.

Given the many fallacies that may not be identified within social work decision-making contexts, it is important to help social work students improve their fallacy recognition skills. In this study, we differentiate between two components of fallacy recognition skills, fallacy assessment and fallacy explanation, which we describe in the following section.

Fallacy Assessment and Fallacy Explanation as Components of Fallacy Recognition Skills

We argue that fallacy recognition skills include two distinct activities which are, however, intertwined to some extent. These activities are (1) fallacy assessment and (2) fallacy explanation.

Fallacy assessment. Fallacy assessment refers to the extent to which a person who analyzes a fallacious argument assesses either the evidence (claims) provided by the argument and/or its conclusion as positive or negative. Let's assume a discussion between social workers regarding the best potential intervention for a given client. Social worker X may state the following: "*We should analyze the client's current situation more thoroughly by using Luhmann's systems theory. Luhmann was an expert in his field and his theory is, to this day, a leading framework in sociology*." Let us assume the following analysis of this argument: "*Claiming that Luhmann is an expert in his field provides no reason to believe that it is justified to use his theory in this context. Reasons for using the theory in this particular case should be provided.*" Here, the analysis implicitly assesses social worker X's evidence (Luhmann was an expert in his field) as well as his conclusion (we should use his theory to analyze the client's situation) as fallacious by rejecting her/his argument (and explaining reasons for her/his rejection).

However, a social worker might simply reject another's argument without further explaining why s/he rejects the argument, for example because s/he assumes implicit claims that justify the rejection. These implicit claims might not be evident for other recipients, yet, if no one objects, the very argument remains rejected. The burden of proof remains with the person who rejected the argument. Thus, it is important to provide a valid and thorough explanation for why a social worker might regard a stated argument as fallacious. We call this activity fallacy explanation.

Fallacy explanation. Fallacy explanation refers to the expression of reasons that explain why a social worker thinks that a previously stated argument is fallacious. For example, a social worker might analyze the argument "*Manualized therapies are the best, because they provide step-by-step guidelines which enhance effectiveness*" simply by answering "*No, I do not think so.*" In this case, no explanation is provided. Alternatively, a social worker might analyze this argument by saying "*No, I do not think so, because the conclusion* (manualized therapies are the best) *and the provided evidence* (they entail step-by-step guidance which enhances effectiveness) *are essentially the same.*" Here, an explanation of the fallacy is provided. This explanation can be considered as reasonable because the very argument violates the *structure* criterion (it implicitly assumes the truth of the conclusion) which is illustrated by this example (Damer, 2011).

Empirical Studies on Fallacy Recognition Skills

Empirical studies on the facilitation of fallacy recognition skills in the field of social work are rare. However, a number of studies investigated such skills in fields other than social work (e.g., Neuman, 2003; Neuman & Weizman, 2003; Ramasamy & Fung, 2015; Tarnoff, 2010; Weinstock et al., 2004).

Using the informal fallacy identification task (IFT; Neuman & Weizman, 2003) with a sample of $N = 202$ K-12 students from an urban high school in Israel, Neuman (2003) found

that about 72% of the students identified the fallacies correctly. Yet, only 16% correctly explained why an argument was fallacious. Similar results (again with K-12 students as participants) were found by Weinstock et al. (2004). In another study, Neuman and Weizman (2003) found that about 40% of ninth-graders correctly identified fallacies. In a second study by Neuman and Weizman (2003), only 25 out of 53 students were successful in identifying fallacies in the IFT.

Studies at the university level also indicated that students often have difficulties with fallacy recognition. For example, in a study with $N = 630$ undergraduate students from various disciplines, Ramasamy and Fung (2015) found an average point score on the ITF of $M = 47.5$ (59.4%), and similar findings were reported by Tarnoff (2010).

In summary, there is a clear gap regarding empirical studies on fallacy recognition skills in social work decision-making contexts. However, empirical evidence from studies in fields other than social work suggest that such skills are often weak among students. Furthermore, previous studies suffered from aspects that can be considered a threat to their internal validity such as failure to counterbalance the order of administered tasks to control for practice effects (Neuman & Weizman, 2003; Ramamsay & Fung, 2015, Weinstock et al., 2004) or the lack of a control group (Neuman, 2003; Ramamsay & Fung, 2015; Tarnoff, 2010; Weinstock et al., 2004). In addition, none of the studies applied a treatment in the context of an experimental design to investigate how to actively foster fallacy recognition skills. Hence, it is important to determine how to best teach fallacy recognition skills to social work students by using experimental designs that compare different kinds of support for students' acquisition of fallacy recognition skills.

Candidate Approaches to Effectively Teach Social Work Students' Fallacy Recognition Skills

As described, there is a lack of empirical studies on how to best teach fallacy recognition skills in social work. Thus, the question how to scaffold social work students' fallacy recognition skills can hardly be answered on an empirical basis. However, there are a number of instructional approaches that could be promising with respect to supporting fallacy recognition skills. Three of them (case-based reasoning, worked examples, and external scripts) will be described in the following.

Case-based Reasoning. Case-based reasoning (CBR) was developed in the 1990's by the group around Janet Kolodner. Learning environments that are based on this approach engage learners in a series of real-world problems (e.g., problem solving in decision-making contexts; see Kolodner, Cox, & Gonzalez-Calero, 2005, for a brief overview of CBR-inspired instructional approaches) in order to motivate them to learn and apply knowledge. In CBR, learning means to extend one's knowledge by aligning new cases with the experience of similar older experiences through (re-)interpretation of both the new and old experiences. Learners reflect on and assess their reasoning in older cases, which makes experiences more applicable for new situations. In doing so, it becomes more likely for learners to recall the most helpful previous experience for a given problem at hand in a new situation (a process which is called *indexing;* Kolodner, 1992). Thus, an important aspect of CBR is the idea that learners adjust and improve their explanations (and thereby also their knowledge) over time (Kolodner, Camp et al., 2003).

Empirical evidence of case-based tools and pedagogical approaches suggests that students who engage in case-based activities perform better in scientific reasoning and peer collaboration than students in more teacher-based learning environments (Kolodner, Camp et al., 2003; Kolodner, Gray, & Fasse, 2003). Other empirical studies also demonstrated the

effectiveness of CBR in facilitating critical thinking skills in various fields other than social work. For example, in a randomized controlled trial with 122 undergraduate nursing students, Hong and Yu (2017) found different styles of case-based instruction to be beneficial regarding effects of critical thinking, measured with a variety of standardized instruments like the California Critical Thinking Skill Test. In a literature review regarding the promotion of critical thinking by using case studies, Popil (2011) concludes that using cases helps develop critical thinking skills for nurses and other health care professionals.

Along with similar approaches such as problem-based learning (e.g., Barrows & Tamblyn, 1980 Hmelo-Silver, 2004) or inquiry learning (Kuhn, Black, Keselman, & Kaplan, 2000), CBR can be labeled as a so-called constructivist instructional approach. Despite their wide-spread use, such approaches have been severely criticized in the past for not being compatible with the human cognitive architecture. For example, Kirschner, Sweller, and Clark (2006) argue that by exposing learners to complex problems that are too difficult for them to solve on their own, their working memory capacity is very likely to be overstrained, causing cognitive overload (for a more extensive description of Cognitive Load Theory, see Sweller, Ayres, & Kalyuga, 2011). Thus, additional scaffolds (a kind of support that helps a learner to learn through activity; Tabak, 2004) that are integrated in CBR and that provide learners with additional guidance might make CBR even more effective.

Worked examples as well as external scripts are two kinds of scaffolds that provide such additional guidance. Both are briefly described in the following.

Worked Examples. In general, worked examples (WE) present step-by-step solutions for problems that are presented to a learner (Sweller et al., 2011). They typically consist of (a) a problem description, (b) a presentation of the different solution steps, and (c) the correct solution (e.g., Wittwer & Renkl, 2010; see Atkinson, Derry, Renkl, & Wortham, 2000 for an overview of instructional principles for WE). WEs point the learner's attention to what is

important regarding the solution of a given problem and reduce cognitive load imposed on working memory during learning (Sweller et al., 2011). As a result, learning should be facilitated, especially when learners use this free capacity to elaborate more deeply on the example (e.g., by self-explaining the single solution steps).

Initially, research on WEs focused on well-structured domains like mathematics (Renkl, 2017). There, studies repeatedly found that learners who studied a series of WEs prior to solving tasks on their own did much better on transfer problem solving tasks than students who – after an initial presentation of the solution procedure – tried to solve problems early-on by themselves (e.g., Carroll, 1994; Schwonke et al., 2009; van Gog, Kester, & Paas, 2011). This robust result has been termed the "worked example effect" (Schwonke et al., 2009; Sweller et al., 2011; van Gog et al., 2011).

More recent studies also showed positive learning effects of WEs in rather ill-structured domains such as mathematics argumentation (Kollar et al., 2014) or medicine (Kopp, Stark, & Fischer, 2008). Kollar et al. (2014) conducted an experimental study with teacher students (N = 101) which revealed that worked examples (compared to problem-solving) could effectively be used to facilitate mathematical argumentation. Kopp et al. (2008) conducted an experimental study with a 2x2 factorial design with two different kinds of WEs (erroneous vs. not erroneous) and feedback (elaborated vs. knowledge of correct result) to facilitate diagnostic knowledge of medical students (N = 158). Their results showed that erroneous examples were more effective regarding the facilitation of diagnostic knowledge than correct examples, especially when they were combined with elaborated feedback.

External Scripts. External scripts (ES) guide learners through a cognitive activity (Kollar, Fischer, & Slotta, 2007). Different from WEs though, they do not show the learner single solution steps together with the finally correct solution, but rather prompt learners to

engage in certain activities themselves while solving the problem. For example, an external script might prompt a learner to analyze an argument by identifying conclusions and claims and subsequently evaluating these conclusions and claims against certain criteria of good argumentation. If the learner subjectively evaluates the effects of applying that activity as successful, repeated practice in this activity may result in the development of an analogous internal script. As a result, it is likely that the learner will transfer his/her knowledge about this activity to future situations.

Empirical evidence for the effectiveness of external scripts comes mainly from research on collaborative learning. A meta-analysis by Vogel, Wecker, Kollar, and Fischer (2017) suggests that collaboration scripts yield considerable positive effects on learners' acquisition of domain-specific knowledge (d = 0.20) and collaboration skills (d = 0.95) compared to unstructured collaboration. For example, in a study by Stegmann, Weinberger, and Fischer (2007) with 120 students of educational science, an ES for the construction of single arguments and an ES for the construction of argument sequences improved students' ability to construct single arguments and argument sequences in a subsequent unguided task and also helped them in their acquisition of knowledge on argumentation. Whether ES have similar effects in social work to help students develop fallacy recognition skills, though, is an open question.

Synergistic Scaffolding

Taken together, WEs and ESs seem to be promising scaffolds to foster the development of social work students' fallacy recognition skills during CBR. However, the identification and explanation of fallacies within social work decision-making contexts may be considered a rather complex skill, as it entails various epistemic activities (Fischer et al., 2014) such as evidence generation, evidence evaluation, and drawing conclusions. To foster such complex practices, Tabak (2004) proposes to combine different scaffolds in order to

create what she calls *synergistic scaffolding.* This suggests that a combination of WE and ES might be more promising regarding the facilitation of social work students' fallacy recognition skills than the presentation of WEs or ESs alone.

Research Questions and Hypotheses

We argued that it is to be expected that social work students' fallacy recognition skills are not fully developed, even though this assumption lacks empirical evidence specifically from the field of social work. Therefore, this study investigates how social work students' fallacy recognition skills can be facilitated. We argued that CBR might be a promising approach in this respect. Yet, its effectiveness might be enhanced by additional instructional guidance, which might be implemented through the integration of WEs and ES. Against this background, this study tried to answer the following research questions:

1. Do social work students have to be trained in social work fallacy recognition skills (i.e., does presentation of additional guidance such as information on arguments, fallacies, and criteria for argumentation yield a learning effect regarding these skills)? We expected learners who received additional guidance prior to CBR to reach higher levels of fallacy recognition skills in a post-test than learners who did not receive additional guidance during CBR (H1).

2. What are the effects of WE and an ES as well as their combination on learners' fallacy recognition skills during CBR (during the intervention)? We expected that both the presence of WE (compared to the absence of WE; H2a) as well as of ES (compared to the absence of ES; H2b) would increase students' fallacy recognition during the intervention (i.e. during the time the two scaffolds were present). Furthermore, we expected a synergistic scaffolding effect such that students who simultaneously received WE and ES would outperform students from all other conditions (H2c).

3. What are the effects of WE and an ES as well as their combination on learners' transfer of fallacy recognition skills? Here again, we expected students who were provided with WE to outperform students on a new case (without guidance) who did not receive WE (H3a), and students who were provided with ES to outperform students who did not receive ES (H3b). Furthermore, and in accordance with RQ2, we expected an interaction effect, namely that students from the combined condition would outperform all other students (H3c).

Methods

Design and Sample

$N = 130$ 6th semester social work students ($M_{age} = 25.3$, $SD_{age} = 6.0$, *women* = 109, *man* = 19, *N/A* = 2) from a University of Applied Sciences (Department of Social Work) in Germany participated in the study. All participants signed informed consent prior to data collection. The sample size required to detect expected medium effect size of Cohen's $f = .25$ at a β – level of .20 was $N = 128$. Participants received financial compensation for their participation. They were randomly assigned to one of the following five conditions: (a) guided CBR with WE and ES ($n = 28$); (b) guided CBR with ES and without WE ($n = 26$); (c) guided CBR with WE and without ES ($n = 25$); (d) guided CBR without ES and without WE ($n = 26$); (e) unguided CBR without ES and without WE ($n = 25$; see Table 1).

(Insert Table 1)

Learning Environment

We created an online web-based learning environment (see Author, 2019 for pictures) which was implemented with Unipark (https://www.unipark.com/). The learning environment consisted mainly of vignettes (Author, 2019) and text boxes in which participants filled in their answers. Each vignette consisted of a social work decision-making scenario (Author, 2019), in which three social workers (X, Y, and Z) exchange fallacious arguments (one after another) regarding the best intervention for a fictitious client. We included five fallacies in

each vignette. Like in real-world practice, the context of the vignettes as well as the fallacies differed slightly (e.g., an isolated veteran with alcohol problems, an 18-year-old boy who lives without his parents and smokes marijuana, a 14-year-old girl stays out long at night and experiences problems in school).

At pre- (time 1) and post-test (time 3), learners were provided with a vignette and one textbox, in which they could fill in their analysis. This was also the case during the intervention (time 2) for participants in conditions with no scaffolds (control and unguided). After the pre-test (time 1), all participants except for the unguided CBR condition were provided with a video (available in German language via Author, 2019) with information regarding arguments and criteria for good argumentation. The video was created based on Damer (2011). It consisted of power point slides and audio explanations regarding how an argument is constructed, how to detect conclusions and claims, what the criteria for good argumentation are, respective examples which were not related to social work in terms of their content, as well as an emphasis on the importance of this knowledge with regard to decision-making contexts in social work.

Experimental Conditions

Unguided CBR condition. The unguided CBR condition took the pre-test (time 1) and analyzed the vignette of the intervention phase (time 2) as well the vignette in the post-test (time 3) without any guidance or scaffolds.

Guided CBR condition. The guided CBR condition (or control condition) was the same as the unguided CBR condition, except that learners saw the video after the pre-test (time 1).

WE condition. During the intervention (time 2), participants in the WE condition were provided with one textbox and an erroneous WE underneath the textbox. The WE was an argument that was not related to social work in terms of its content. It was erroneous in the

sense that it provided a conclusion, a valid claim, and five additional claims that each violated one of the five criteria for good argumentation. An example for such an argument would be "*Riding a motor bike is dangerous* (conclusion), *because motor bikes have no safety zone* (valid argument) *and because driving a car is less dangerous* (claim that violates the relevance criterion)."

ES condition. The ES condition was provided with parts of the vignette in 3 stages for each social worker's argument. Participants first read social worker X's argument and were asked to identify X's overall conclusion and type it in a textbox below the argument (stage 1). On the following screen, participants were provided with the conclusion they previously identified as well as with ten textboxes. They were asked to identify all claims in X's argument and to type each claim in a separate textbox (ten was the limit; not all textboxes had to be used; stage 2). On the next screen, participants were provided with both the conclusion they identified in stage 1, as well as all the claims that they identified in stage 2. They were asked to analyze whether one of the claims they identified, resp. the argument of social worker X, violated one of the five criteria. If this was the case, they were asked to type an explanation of how it violated the criterion into a respective text box (each textbox had a headline referring to one of the criteria; stage 3). After they completed stage 3, the procedure repeated for social worker Y's argument and then for social worker Z's argument.

WS + ES condition. Likewise, participants in the condition with WE and ES were provided with the same learning environment as the ES condition during time 2. However, they were additionally provided with the part of the WE that was relevant for a specific textbox (i.e., under the textbox in which they should analyze whether an argument violated the criteria *structure*, they saw the part of the WE that violated this criterion).

Procedure

Data collection was conducted in March and April 2018. Each participant worked on a laptop with the digital learning environment. The average time-on-task (for the whole process from pre- to post-test) was 86.3 minutes (*SD* = 18.95, range: 46.75 to 149.73) while there was no time limit. During pre-test, learners completed a questionnaire regarding demographic data, followed by a multiple-choice test to determine their domain-specific prior knowledge of social work theories and methods. Then, students in all conditions were asked to analyze a vignette without guidance, serving as a pre-test to assess their fallacy recognition skills (time 1). Afterwards, students, except those in the *unguided* condition, were provided with the video. Subsequently, all groups had to analyze another vignette, depending on their experimental condition either with WE or ES, both or none of these scaffolds (time 2). After this analysis, all participants had to analyze a third vignette as a post-test (time 3). Participants of the *unguided* condition saw the video after the post-test for ethical reasons and they were offered to analyze a fourth vignette (time 4) on a voluntary basis (17 of the initial 25 participants of the *unguided* condition completed this extra vignette). The fourth vignette, however, was not included in the analyses. Table 2 shows an overview of the procedure.

(Insert Table 2)

Dependent Variables and Coding

We were interested in two dependent variables (DV), namely students' fallacy assessment (FA) and fallacy explanation (FE) skills. These were measured via an analysis of participants' analyses of the vignettes (at pre-test, during intervention, and at post-test). Each vignette included five fallacies that were created based on Damer (2011) and Gambrill and Gibbs (2017). Unlike the vignette of the treatment phase (time 2), vignettes of the pre- (time 1) and post-test (time 2) were counterbalanced to control for task difficulty. We developed and iteratively refined a coding scheme to code learners' analyses regarding our variables of

interest. Two independent coders coded FA and FE for > 10% of the data for each fallacy. Coders were blind to whether data came from pre- (time 1) or post-test (time 3). During the intervention phase, the structure of the data from the ES conditions made it obvious that these learners had an ES available or not (i.e., the structure of these answers resembled the prompts of the ES to list the conclusion, all claims and violations of the criteria for good argumentation). Thus, coders were not blind with respect to this condition; yet, they were blind to whether students had worked with or without WE. Interrater agreement was determined by the intra-class coefficient (ICC; two-way mixed, absolute, single rater) and an ICC of .70 was deemed sufficient (Koo & Li, 2016). If an ICC was not sufficient, all cases which were rated differently were discussed and the coding scheme was refined respectively (see Author, 2019 for the protocol). Then, the coders coded another > 10% of the data. We continued with this process until a sufficient ICC was achieved for all 30 variables (three vignettes, each containing five fallacies, each of which were coded with respect to the two variables FA and FE). Afterwards, all remaining data was coded by the first author. The coding scheme differed with respect to our two dependent variables of interest – fallacy assessment and fallacy explanation – as described in the following section.

Fallacy assessment. The variable fallacy assessment refers to the extent to which participants were able to identify fallacies. FA was coded for each fallacy on a 4-point Likert scale with the options *not analyzed* (0 points), *not fallacious* (1 point), *so-so* (2 points) and *fallacious* (3 points). Since five fallacies were included in each vignette, the scoring range was 0 to 15 points for the pre-test, the intervention, and the post-test each. Regarding FA, participants answers were coded as *not analyzed* for a particular fallacy, if the fallacy was not addressed in their answer at all. The code *not fallacious* was given when claims and/or the conclusion of a fallacy were incorrectly assessed as positive. The code *so-so* was given when the analysis assessed a fallacy in a vague manner or to some extent positive and to some

extent negative at the same time, or when the conclusion of a fallacious argument was assessed positively, while the respective evidence was assessed negatively and also if the conclusion of a fallacious argument was assessed negatively, without explicitly addressing the respective fallacy. Answers were coded as *fallacious* when claims and/or the conclusion of a fallacy were assessed negatively. The range of the achieved ICCs for the 15 variables regarding FA was .71 to 1.0.

Fallacy explanation. The variable fallacy explanation refers to the quality of an explanation regarding the question why an argument is fallacious. FE was coded on a 4-point Likert scale with the options *N/A* (0 points), *no explanation* (0 points), *low quality* (1 point), *medium quality* (2 points) and *high quality* (3 points) with the latter three referring to the quality of the explanation. Since five fallacies were included in each vignette, the scoring range was 0 to 15 points, each with time 1, 2 and 3. To code FE, we developed a set of two 'ideal' explanations for each fallacy. Where possible, this was done deductively by Damer's (2011) theory. Otherwise, we developed explanations inductively by examining the data and deciding which explanations were the most comprehensible ones. FE was coded with *N/A* when FA for the same fallacy was coded either with *not analyzed* or *not fallacious.* FE was coded with *low* when some kind of explanation for why a certain fallacy is fallacious was provided, and when this explanation did not include one of the two 'ideal' explanations. FE was coded with *medium* when the analysis included one (or a specific, depending on the fallacy) explanation of the respective set of 'ideal' explanations. The code *high* was given when the analysis included two (or a specific, depending on the fallacy) explanation/s of the respective set (see Author, 2019 for all sets of 'ideal' explanations). The range of the achieved ICCs for the 15 variables regarding FE was .71 to 1.0 (see Author, 2019 for a detailed table).

Control Variables and Instruments

Domain-specific prior knowledge. All participants completed a multiple-choice test (Author, 2019) on domain-specific prior knowledge that focused on social work theories and methods. In its original version, the test consisted of 40 questions with four answer opportunities each. A different number of answers (either one, two, three or four) could be correct for a specific question. This initial version was tested in a pilot test with nine bachelor of social work students and neither ceiling nor floor effects were observed. Students received one point for each correctly identified true or false answer. An initial reliability analysis revealed an insufficient Cronbach's Alpha of $\alpha = .43$. Thus, to arrive at a more reliable measure, items with low item-scale correlations were removed. The final test consisted of 27 items with a sufficient reliability of $\alpha = .60$ (Cronbach'a Alpha). Correlations between domain-specific knowledge and the DVs at the post-test were significant but rather small (Pearson's $r = .30$, $p = .001$ for FA and .32, $p = .003$ for FE). Even though differences between conditions regarding domain-specific prior knowledge were small and non-significant (all Fs $(1,101) < 1$, *n.s.*), it was included as a covariate in the analyses regarding our research questions.

Time-on-task. Time-on-task was recorded automatically by the learning environment for the whole data collection process. Preliminary analyses showed that students in the ES conditions displayed a significantly higher time-on-task than students from conditions in which no ES was present (F (1, 101) = 6.102, $p = .015$, *partial* $\eta^2 = .06$ (all other effects: Fs $(1,101) > 1$, *n.s.*). Correlations between time-on-task and the DVs were small though ($r = .10$, $p = .35$ for FA and $r = .09$, $p = .38$ for FE). Still, when analyzing the data regarding our research questions, we included this variable as a covariate.

Statistical Analyses

In a preliminary analysis, we carried out paired-samples t-tests to determine if learners increased their fallacy recognition skills over time (i.e., from pre- to post-test), independent from experimental conditions, differences in prior knowledge, or time-on-task.

To answer RQ 1 (effects of guided vs. unguided CBR), we ran one-way independent ANCOVAs with FA or FE at post-test (time 3) as DVs, guided (all four conditions combined) vs. unguided CBR as a between-subject factor, and the pre-test (time 1; respectively for FA or FE), prior knowledge, and time-on-task as covariates.

The correlation between the two DVs was substantial ($r = .83$, $p < .001$) at post-test, therefore we assumed multicollinearity and decided not to carry out a MANCOVA to address RQs 2 and 3.

Instead, to answer RQ 2 (effects of WE and ES during the intervention) and RQ 3 (effects of WE and ES at post-test) we applied two-way independent ANCOVAs with FA and FE at time 2 for RQ 2 and at time 3 for RQ 3 as DVs, WE and ES as between-subject factors, and the pre-test (time 1; respectively for FA or FE), prior knowledge, and time-on-task as covariates.

The α-level for all statistical analyses was set to 0.05. *Partial* η^2 was used as a measure of effect size. Values of about .01 were considered as a small effect, values of about .06 as a moderate effect and values of about .14 as a large effect (Cohen, 2013). SPSS syntax is available via Author (2019).

Results

With respect to ANCOVAs, Shapiro-Wilk tests indicated that the assumption of normality was violated for the variable FA in the ES condition ($p = .038$) and the WE+ES condition ($p = .004$) as well as for FE in the WE+ES condition ($p = .025$) during guided CBR (time 2). The same was true for FE in the WE condition ($p = .040$), as well as for the variable

FA in the ES condition (p = .034) at post-test (time 3). However, this was considered unproblematic since group sizes were equal and histograms indicated mesokurtic distributions (Quinn & Keough, 2002). Furthermore, a violation of the normality assumption is only considered potentially problematic if it results in variance heterogeneity, which was not the case for our data as indicated by Levene's tests which were all non-significant (Quinn & Keough, 2002).

We encountered a few outliers based on the IQR-method. Since there was no theoretical background to justify their exclusion on the one hand and no outliers were found based on the z-score method (no values less than -3 or greater than 3), outliers were not excluded.

Table 3 provides an overview of the means and standard deviations of students' scores on the prior knowledge test, on FA and FE in time 1, 2, and 3, and their time-on-task.

(Insert Table 3)

Preliminary Analysis: Did Learners Learn over Time?

Paired-samples t-tests yielded a statistically significant increase in students' FA score from time 1 (M = 6.91, SD = 3.24) to time 3 (M = 7.80, SD = 3.32), t (125) -2.487, p = .007, d = 0.27 (one-tailed). The mean increase was -0.89 (95% *CI*: -1.596, -0.181; negative values due to direction of comparison which was time 1 minus time 3). Likewise, we found a statistically significant increase in students' FE scores from time 1 (M = 3.53, SD = 2.11) to time 3 (M = 4.27, SD = 2.69), t (125) -2.628, p = .005, d = 0.31 (one-tailed). The mean increase was -0.74 (95% *CI*: -1.294, -0.182).

Research Question 1: What are the Effects of Guided vs. Unguided CBR?

One-way independent ANCOVAs were conducted to test the effect of guided CBR (all four conditions in which the video with information on arguments, fallacies, and criteria

for good argumentation was presented before treatment) vs. unguided CBR (the condition that neither received the video nor further scaffolds) on FA and FE at post-test.

Regarding FA, guided learners performed better (M = 8.01, SD = 3.10) than unguided learners (M = 6.62, SD = 3.32). This difference was statistically significant (F (1, 115) = 4.60, p = .034, *partial* η^2 = .038).

With regards to FE, guided learners (M = 4.33, SD = 2.69) outperformed unguided learners (M = 3.46, SD = 2.43). This improvement was marginally significant F (1, 115) = 2.675, p = .105, *partial* η^2 = .023).

Therefore, H1 was partially corroborated.

Research Question 2: What are the Effects of WE and an ES as well as their Combination on Learners' Fallacy Recognition Skills during Guided CBR (during the Intervention)?

Two-way independent ANCOVAs were conducted to explore the impact of WE, ES, and their combination on social work students' fallacy recognition skills (FA and FE) during guided CBR (i.e., during the application of the scaffolds). The unguided condition was excluded from these analyses.

In terms of FA, learners in the WE group (M = 9.14, SD = 2.83) performed better than those in the ES condition (M = 8.64, SD = 2.33). The combined condition (M = 7.75, SD = 2.14) and the control condition (M = 7.72, SD = 2.69) were similar in their performance and both inferior compared to the WE and ES condition. However, there was a non-significant main effect for WE (F (1, 93) = .214, p = .645), a non-significant main effect for ES (F (1, 93) = .497, p = .483), but a significant disordinal interaction effect between WE and ES (F (1, 93) = 4.080, p = .046, *partial* η^2 = .042). This indicates that both scaffolds were effective as long as they were presented without the other. Once the scaffolds were combined, however, students' FA was no better than in the control condition.

As for FE, learners in the WE condition (M = 5.45, SD = 1.99) and in the ES condition (M = 5.40, SD = 1.92) performed better than those in the combined condition (M = 4.86, SD = 1.38) and in the control condition (M = 4.32, SD = 2.06). Yet, we found a non-significant main effect for WE (F (1, 93) = 0.459, p = .500), a non-significant main effect for ES (F (1, 93) = .001, p = .978), but again a significant disordinal interaction effect between WE and ES (F (1, 93) = 4.672, p = .033, *partial* η^2 = .048). Just like for FE, this indicated positive effects of each of the two scaffolds when presented without the other. Yet, a combination of the two scaffolds seems to be no better than CBR without ES and WE. Therefore, H2a, H2b, and H2c were rejected.

Research Question 3: What are the Effects of Worked Examples and External Scripts as well as their Combination on Social Work Students' Fallacy Recognition Skills in Post-test?

Two-way independent ANCOVAs were conducted to explore the impact of WE, ES, and their combination on social work students' fallacy recognition skills (FA and FE) during post-test.

Regarding FA, learners in the WE condition (M = 8.91, SD = 2.67) performed best, and, surprisingly, learners in the control condition (M = 8.17, SD = 3.74) second best. Learners in the combined condition (M = 7.65, SD = 3.32) and those in the ES condition (M = 7.42, SD = 3.43) performed similarly. There was a non-significant main effect for WE (F (1, 89) = .468, p = .496), a marginally significant negative main effect for ES (F (1, 89) = 2.827, p = .083, *partial* η^2 = .031), indicating a detrimental effect of ES on FA compared to no scaffold, and a non-significant interaction effect between WE and ES (F (1, 89) = .108, p = .743).

Concerning FE, again surprisingly, learners in the control condition (M = 4.83, SD = 3.23) performed best, followed by learners in the WE condition (M = 4.77, SD = 2.02).

Learners in the combined condition (M = 4.04, SD = 2.96) and in the ES condition (M = 3.75, SD = 2.31performed worse than the other two groups. We observed a non-significant main effect for WE (F (1, 89) = .020, p = .888), a significant negative main effect for ES (F (1, 89) = 3.992, p = .049, *partial* η^2 = .043), again, suggesting that ES have a negative effect on FE compared to no intervention, and a non-significant interaction effect between WE and ES (F (1, 89) = .020, p = .926). Therefore, H3a, H3b, and H3c were rejected.

Discussion

The aim of this article was to examine the effects of an online CBR environment on social work students' fallacy recognition skills. We further were interested in the effects of different versions of the CBR environment that differentiated with respect to whether it included or excluded two different kinds of scaffolds: Worked examples and external scripts.

On average, all learners improved their fallacy recognition skills. Even though effect sizes were small, the average increase of about 1 point (range: 0 to 15 points) from pre- to post-test for fallacy assessment as well as for fallacy explanation might well be regarded as practically relevant, considering the complexity of these skills in combination with the relatively short intervention. Thus, CBR can be considered a useful approach to facilitate social work students' fallacy recognition skills, especially when applied over a longer period such as a whole semester. This might allow learners to benefit even more from CBR, due to its focus on learning from experience (Schank, 1999). Further research is needed to test this assumption.

Guided CBR seems to be favorable in comparison to unguided CBR. Taking the content of our provided guidance into account, it seems reasonable to consider Damer's (2011) theory as promising for the facilitation of fallacy recognition skills.

During the intervention (time 2), students in the ES-only and the WE-only condition showed higher values on FA and FE than students in the control condition. The significant

disordinal interaction effect indicates that WE as well as ES are effective when presented in isolation, but that their support decreases when they are combined. This means that the scaffolds are not effective in general, but only when they are not combined with each other. On the one hand, our results are in line with evidence from studies that showed positive effects of WE (e.g., Renkl, 2017) and ES (e.g., Vogel et al., 2017) on learning processes. Yet, we found no evidence for synergistic scaffolding (Tabak, 2004): To the contrary, the results indicate that the two scaffolds lose their effectiveness once they are combined with the other. This result might be explained by the necessity for learners to switch their attention between the two scaffolds back and forth, which might have been a strain on learners' working memory, leaving them with little cognitive capacity to focus on the actual problem solving processes (*split attention effect;* Kollar et al., 2014; Sweller et al., 2011). Yet, further research is needed to test this interpretation.

The positive results for the ES-only and the WE-only conditions could not be maintained as indicated by the post-test (time 3). It is however not unusual that learners need some time to transfer new learning strategies to benefit from them. Empirical evidence suggests that learners can perform significantly better in delayed tests compared to tests that are administered immediately after the treatment. This has been shown for ES (e.g., Lee, 2018) as well as for erroneous WE (e.g., McLaren, Adams, & Mayer, 2015). Therefore, future studies should use delayed tests to investigate potential long-term effects of ES and WE on social work students' fallacy recognition skills. Furthermore, considering the complexity of fallacy recognition skills, longer interventions might contribute to the effectiveness of ES and WE on these skills.

Another possible explanation for the rather disappointing outcomes is that the particular design of the two scaffolds might have been suboptimal. With respect to the ES, maybe its design was too far from students' prior skill level, thus failing to provide them with

a Zone of Proximal Development (Vygotsky, 1980) that they could productively use to further develop their fallacy recognition skills. Given the complexity of fallacy recognition skills, this might indeed have been the case. Future research should thus seek to identify designs of ES that social work students can more easily take up when analyzing cases. The fact that ES even had a negative effect makes it even more necessary to develop further alternatives for the design of ES to support students' fallacy recognition skills. WEs, in turn, might become more effective if they show how exactly the fallacy that students are confronted with can be detected and resolved, rather than illustrating the concepts of fallacy analysis with a different fallacy. From this perspective, our WEs might not have been rich enough (Renkl, 2014). Also, research by Schworm and Renkl (2007) shows that with respect to fostering skills, WEs seem to be only effective when they are combined with self-explanation prompts, which were not presented in our study. Thus, future research should try to explore how to design scaffolds that more closely relate to social work students' prerequisites regarding fallacy recognition.

Limitations and Conclusions

As with all research, the results of our study must be considered in the light of several limitations. First, our study might have been underpowered to yield effects of WE and ES on fallacy recognition skills, since estimating the expected effect size was difficult due to the absence of previous studies in social work. Expected effect sizes may be much smaller, and thus a larger sample would be required to be able to show these effects. Second, we applied a convenience non-probability sampling method which limits the external validity of our results (Nugent, 2010). Third, we have chosen the fallacies used in our vignettes based on conceptual literature, assuming that these fallacies are rather prevalent in social work practice. However, no further systematic process guided our selection of the fallacies. Fourth, our worked examples were not very rich in that they did not resemble *the steps* to a solution

but rather only the solution itself. Fifth, the intervention was rather short, considering the complexity of fallacy recognition skills.

Despite these limitations, we believe that our study is among the first to provide empirical evidence regarding fallacy recognition skills of social work students. Our evidence suggests that social work students need support to identify fallacies and even more so to provide valid reasons to explain why a fallacious argument is fallacious. CBR seems promising in this regard and could be easily implemented in social work education with a digital learning environment such as the one that was used in this study, if devices with web-access are available for learners. For example, this might be done in the form of a homework assignment or as part of a video lecture.

However, our results provide only limited evidence that worked examples and external scripts can further boost the effectiveness of CBR. We argue that further systematic research is needed to identify which features in the design of WE and ES can be improved to make them more effective.

Disclosure Statement

The authors do not have any conflicts of interest.

Funding

The research presented in this work was supported by the Elite Network of Bavaria [Project number: K-GS-2012-209].

References

Author. (2019). Effects of worked examples and external scripts on social work students' fallacy revelation skills. Open Science Framework. Retrieved from osf.io/5p86x, offline for peer-review

Aglen, B. (2016). Pedagogical strategies to teach bachelor students evidence-based practice: A systematic review. *Nurse Education Today*, *36*, 255–263. https://doi.org/10.1016/j.nedt.2015.08.025

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research*, *70*, 181–214. https://doi.org/10.3102/00346543070002181

Barrows, H. S., & Tamblyn, R. M. (1980). *Problem-based learning: An approach to medical education*. *Springer series on medical education: v. 1*. New York: Springer Pub. Co.

Carroll, W. M. (1994). Using worked examples as an instructional support in the algebra classroom. *Journal of Educational Psychology*, *86*, 360–367. https://doi.org/10.1037/0022-0663.86.3.360

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hoboken: Taylor and Francis.

Damer, T. E. (2011). *Attacking faulty reasoning: A practical guide to fallacy-free arguments* (7. ed.). Boston, MA: Wadsworth Cengage Learning.

Fischer, F., Kollar, K., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., . . . Eberle, J. (2014). Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda in education. *Frontline Learning Research*, 28–45.

Gambrill, E. D. (2011). *Propaganda in the helping professions*. Oxford: Oxford University Press USA.

Gambrill, E. D. (2013). *Social work practice: A critical thinker's guide* (3rd ed.). New York: Oxford University Press.

Gambrill, E. D. (2019). *Critical thinking and the process of evidence-based practice*. New York, NY: Oxford University Press.

Gambrill, E. D., & Gibbs, L. (2017). *Critical thinking for helping professionals: A skills-based workbook*. New York, NY: Oxford University Press.

Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review*, *16*, 235–266. https://doi.org/10.1023/B:EDPR.0000034022.16470.f3

Hong, S., & Yu, P. (2017). Comparison of the effectiveness of two styles of case-based learning implemented in lectures for developing nursing students' critical thinking ability: A randomized controlled trial. *International Journal of Nursing Studies*, *68*, 16–24. https://doi.org/10.1016/j.ijnurstu.2016.12.008

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*, 75–86. https://doi.org/10.1207/s15326985ep4102_1

Kollar, I., Fischer, F., & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning and Instruction*, *17*, 708–721.

Kollar, I., Ufer, S., Reichersdorfer, E., Vogel, F., Fischer, F., & Reiss, K. (2014). Effects of collaboration scripts and heuristic worked examples on the acquisition of mathematical

argumentation skills of teacher students with different levels of prior achievement. *Learning and Instruction*, *32*, 22–36. https://doi.org/10.1016/j.learninstruc.2014.01.003

Kolodner, J. L. (1992). An introduction to case-based reasoning. *Artificial Intelligence Review*, *6*, 3–34. https://doi.org/10.1007/BF00155578

Kolodner, J. L., Camp, P. J., Crismond, D., Fasse, B., Gray, J., Holbrook, J., . . . Ryan, M. (2003). Problem-based learning meets case-based reasoning in the middle-school science classroom: Putting learning by design™ into practice. *Journal of the Learning Sciences*, *12*, 495–547. https://doi.org/10.1207/S15327809JLS1204_2

Kolodner, J. L., Cox, M. T., & Gonzalez-Calero, P. A. (2005). Case-based reasoning-inspired approaches to education. *The Knowledge Engineering Review*, *20*, 299–303. https://doi.org/10.1017/S0269888906000634

Kolodner, J. L., Gray, J., & Fasse, B. B. (2003). Promoting transfer through case-based reasoning: rituals and practices in learning by design classro. *Cognitive Science Quaterly*, *3*, 183–232.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*, 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kopp, V., Stark, R., & Fischer, M. R. (2008). Fostering diagnostic knowledge through computer-supported, case-based worked examples: effects of erroneous examples and feedback. *Medical Education*, *42*, 823–829. https://doi.org/10.1111/j.1365-2923.2008.03122.x

Kuhn, D., Black, J., Keselman, A., & Kaplan, D. (2000). The development of cognitive skills to support inquiry learning. *Cognition and Instruction*, *18*, 495–523. https://doi.org/10.1207/S1532690XCI1804_3

Lee, Y. H. (2018). Scripting to enhance university students' critical thinking in flipped learning: Implications of the delayed effect on science reading literacy. *Interactive Learning Environments*, *26*, 569–582. https://doi.org/10.1080/10494820.2017.1372483

McLaren, B. M., Adams, D. M., & Mayer, R. E. (2015). Delayed learning effects with erroneous examples: a study of learning decimals with a web-based tutor. *International Journal of Artificial Intelligence in Education*, *25*, 520–542. https://doi.org/10.1007/s40593-015-0064-x

Neuman, Y. (2003). Go ahead, prove that God does not exist! On high school students' ability to deal with fallacious arguments. *Learning and Instruction*, *13*, 367–380. https://doi.org/10.1016/S0959-4752(02)00011-7

Neuman, Y., & Weizman, E. (2003). The role of text representation in students' ability to identify fallacious arguments. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *56*, 849–864. https://doi.org/10.1080/02724980244000666

Nugent, W. R. (2010). Probability and sampling. In B. A. Thyer (Ed.), *The handbook of social work research methods* (2nd ed., pp. 37–50). Los Angeles: SAGE.

Parrish, D. E., & Rubin, A. (2012). Social workers' orientations toward the evidence-based practice process: a comparison with psychologists and licensed marriage and family therapists. *Social Work*, *57*, 201–210.

Pope, N. D., Rollins, L., Chaumba, J., & Risler, E. (2011). Evidence-based practice knowledge and utilization among social workers. *Journal of Evidence-Based Social Work*, *8*, 349–368. https://doi.org/10.1080/15433710903269149

Popil, I. (2011). Promotion of critical thinking by using case studies as teaching method. *Nurse Education Today*, *31*, 204–207. https://doi.org/10.1016/j.nedt.2010.06.002

Quinn, G. P., & Keough, M. J. (2002). *Experimental design and data analysis for biologists*. Cambridge: Univ. Press.

Ramasamy, S., & Fung, H. P. (2015, October). *The analysis of informal reasoning fallacy and critical thinking disposition among malaysian undergraduates*. Proceedings of the Aeu International Research Conference, Kuala Lumpur.

Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, *38*, 1–37. https://doi.org/10.1111/cogs.12086

Renkl, A. (2017). Instruction based on examples. In R. E. Mayer & P. A. Alexander (Eds.), *Educational psychology handbook series. Handbook of research on learning and instruction* (pp. 325–348). New York, NY, London: Routledge.

Schank, R. C. (1999). Dynamic memory revisited. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511527920

Schwonke, R., Renkl, A., Krieg, C., Wittwer, J., Aleven, V., & Salden, R. (2009). The worked-example effect: Not an artefact of lousy control conditions. *Computers in Human Behavior*, *25*, 258–266. https://doi.org/10.1016/j.chb.2008.12.011

Schworm, S., & Renkl, A. (2007). Learning argumentation skills through the use of prompts for self-explaining examples. *Journal of Educational Psychology*, *99*, 285–296. https://doi.org/10.1037/0022-0663.99.2.285

Stegmann, K., Weinberger, A., & Fischer, F. (2007). Facilitating argumentative knowledge construction with computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning*, *2*, 421–447. https://doi.org/10.1007/s11412-007-9028-y

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory* (1st ed.). *Explorations in the Learning Sciences, Instructional Systems and Performance Technologies: Vol. 1*. New York, NY: Springer Science+Business Media LLC.

Tabak, I. (2004). Synergy: A complement to emerging patterns of distributed scaffolding. *Journal of the Learning Sciences*, *13*, 305–335. https://doi.org/10.1207/s15327809jls1303_3

Tarnoff, J. (2010). *An investigation of the role of confirmation bias in the evaluation of informal reasoning fallacies*. Philadelphia.

Van der Zwet, R. J. M., Kolmer, D. M. B. G., & Schalk, R. (2016). Social workers' orientation toward the evidence-based practice process: A dutch survey. *Research on Social Work Practice*, *26*, 712–722. https://doi.org/10.1177/1049731514540340

Van Gog, T., Kester, L., & Paas, F. (2011). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology*, *36*, 212–218. https://doi.org/10.1016/j.cedpsych.2010.10.004

Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-cognitive scaffolding with computer-supported collaboration scripts: A meta-analysis. *Educational Psychology Review*, *29*, 477–511. https://doi.org/10.1007/s10648-016-9361-7

Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes*: Harvard University Press.

Weinstock, M., Neuman, Y., & Tabak, I. (2004). Missing the point or missing the norms? Epistemological norms as predictors of students' ability to identify fallacious arguments. *Contemporary Educational Psychology*, *29*, 77–94. https://doi.org/10.1016/S0361-476X(03)00024-9

Wittwer, J., & Renkl, A. (2010). How effective are instructional explanations in example-based learning? A meta-analytic review. *Educational Psychology Review*, *22*, 393–409. https://doi.org/10.1007/s10648-010-9136-5

Table 1.

Design

		Guided		Unguided
		Worked Examples		
		With	Without	
External Scripts	With	*n* = 28	*n* = 26	
	Without	*n* = 25	*n* = 26	
				n = 25

Table 2

Overview of the Procedure

Procedure phase	Groups	Variables
Pre-Test (time 1)	all conditions	Demographic variables Domain-specific prior knowledge (social work), FA and FE
Video	Only guided conditions	
Treatment (time 2)	all conditions	FA and FE
Post-Test (time 3)	all conditions	FA and FE
Video	unguided	
Optional test (time 4)	unguided	FA and FE

Table 3

Mean Frequencies and Standard Deviations of the Scores for the two Dependent Variables Fallacy Assessment (FA) and Fallacy Explanation (FE) with Regards to their Measurement Points and the Experimental Conditions

Construct	Group									
	Unguided		Control		WE		ES		WE + ES	
	M	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*
DSPK	21.24	1.59	21.08	(2.45)	20.72	(3.18)	21.00	(2.51)	21.50	(2.85)
Time-on-task (min)	74.37	12.64	79.10	(15.17)	83.88	(20.39)	90.62	(17.89)	90.78	(20.33)
FA (time 1)	7.36	(2.70)	6.65	(3.71)	7.28	(3.32)	7.15	(3.11)	6.00	(3.46)
FA (time 2)	7.12	(2.78)	7.88	(2.76)	9.40	(2.77)	8.77	(2.37)	7.75	(2.14)
FA (time 3)	6.68	(3.05)	8.32	(3.74)	9.00	(2.84)	7.36	(3.37)	7.65	(3.32)
FE (time 1)	4.10	(1.85)	3.65	(2.55)	3.40	(2.00)	3.50	(2.14)	3.04	(2.05)
FE (time 2)	3.00	(1.53)	4.46	(2.14)	5.72	(2.05)	5.42	(1.88)	4.86	(1.38)
FE (time 3)	3.60	(2.48)	4.96	(3.22)	4.92	(2.25)	3.84	(2.30)	4.04	(2.96)

Note. DSPK = domain-specific prior knowledge, WE = worked example, ES = external script, FA = fallacy assessment, FE = fallacy explanation

Part III

General Discussion, Implications and Conclusion[9]

[9] This part functions as complement to Part I („ausführliche Zusammenfassung“).

8. General Discussion

The following chapter provides a brief reflection on the goals of this dissertation (chapter 8.1), followed by a general discussion of its findings (chapter 8.2).

8.1 Discussion of addressed Problems and the Goals of the Dissertation

The overarching problem on which this thesis is based on is the scarce use of Evidence-based Practice in social work practice, namely of its two major approaches, the process of Evidence-based Practice (EBP) and Empirically Supported Treatments (ESTs). Suboptimal training of these concepts in social work education was identified as one important reason.

Therefore, the three main goals of the first empirical study (chapter 6) were (1) to provide an overview of educational interventions with the aim to foster the use of these approaches, (2) to investigate how these interventions were conceptualized as well as (3) to assess the quality of the related studies in order to find out how to best teach evidence-based practice to social work students and practitioners. A systematic review was chosen as a methodological approach to tackle the questions related to these goals.

The second study (chapter 7) is an experimental study on the facilitation of social work students' fallacy recognition skills. These skills – specifically the skills to identify as well as to explain fallacies – are considered prerequisites for the motivation to engage in as well as for the use of EBP (Aglen, 2016; Gambrill, 2019; Profetto-McGrath, 2005). The main goal of the second study was to investigate ways to effectively facilitate these skills. An instructional approach (Case-based Reasoning; CBR) as well as two additional scaffolds (worked examples and an external script) were introduced as promising means to foster fallacy recognition skills. Accordingly, the second study tested the effects of

worked examples and an external script in a case-based digital learning environment on social work students' fallacy recognition skills.

The following chapter reports and discusses the results of the two empirical studies.

8.2 General Discussion of Findings

The first empirical study (chapter 6) aimed to provide an overview of studies that investigate the facilitation of EBP and/or the use of ESTs in social work. Further objectives were to systematically evaluate the respective body of research regarding the question how to best teach EBP and/or the use of ESTs to social work students and practitioners. A systematic literature search revealed 27 eligible studies (13 on EBP, 12 on one or more EST/s and two on both approaches). These studies were evaluated with respect to their (1) instructional approach, their reported (2) effects as well as their (3) methodological quality.

A guided student-centered approach was the most prevalent *instructional approach* among the studies. This is in line with the calls from many social work scholars to apply Problem-based Learning in EBP education (e.g., Drake, Hovmand, Jonson-Reid, & Zayas, 2007; Drisko & Grady, 2018; Franklin, 2007; Gambrill, 1999, 2007).

Guided student-centered instruction seems to be beneficial as reflected by the many positive *effects* (89%) of the studies included in the review. However, this finding is rather inconclusive for two reasons. First, other approaches such as more teacher-centered ones also yielded almost solely positive results. A second reason refers to the high prevalence of one-group-designs (see also the discussion of the studies' quality below). An intentional attempt to foster the acquisition of knowledge is in general likely to be successful *to some extent*. Therefore, it is of limited value to know if a particular

teaching approach yields positive results at all. The more important question is which approaches work best under what circumstances and in comparison to other approaches. For example, empirical evidence from psychological research suggests that novices with little prior knowledge of the subject matter may benefit more from rather teacher-centered approaches such as Direct Instruction (Slavin, 2018), while more experienced learners might benefit more from rather student-centered approaches such as Problem-based Learning (Hmelo-Silver, 2004; Kollar & Fischer, 2019). Thus, the general call for Problem-based Learning that is made in many conceptual articles with regard to EBP education in social work (e.g., Drake et al., 2007; Drisko & Grady, 2018; Franklin, 2007; Gambrill, 1999, 2007) might at least be questionable under certain circumstances. The findings of the review cannot empirically support the assumption that student-centered approaches are the best way to facilitate EBP in social work under any circumstances and in comparison to any other instructional approach.

Concerning the studies' *quality*, four aspects are especially important: (a) The small number of controlled designs, (b) the low/or respectively the lack of reliability of the applied measure instruments, (c) the low validity of many studies and the ratings of (d) the studies' risk of bias.

First, only six out of the 27 studies used a *controlled design* (3 experimental and 3 quasi-experimental studies that mainly applied a passive control group or treatment as usual as a control condition). This provides empirical evidence for Yaffe's[10] (2013) assumption that experimental studies are underrepresented in research on social work education. This finding is also in line with empirical evidence that suggests that most of the social work research is descriptive or explorative in its nature (Rosen, Proctor, &

[10] At the time of writing this doctoral thesis, Joanne Yaffe is the editor-in-chief of the *Journal of Social Work Education*.

Staudt, 1999). One reason for this might be a respective research tradition in social work that relates to the idea that experimental designs are not capable of appropriately investigating complex social problems (Gitterman & Knight, 2013; Webb, 2001) and that such empirical research lacks a focus on the individual by concentrating on the investigation of general patterns (Micheel, 2010). The question to which extent such designs are reasonably applicable regarding questions of effectiveness of interventions in social work practice is not subject to this doctoral thesis. However, the scarce amount of experimental designs is problematic in the context of research on social work education, considering the value of controlled designs for the improvement of educational practice (e.g., Bruer, 1994; McGilly, 1998). Furthermore, and strictly spoken, only randomized controlled studies allow strong causal inferences (Shadish, Cook, & Campbell, 2002).

A second problem regarding the studies' quality refers to the way with which data was collected and analyzed. Few studies reported information on *reliability* (e.g., only 13% of the quantitative studies reported on test-retest reliability) of their measure instruments. Also, it should be noted that in some cases in which such information was provided, part of it gave also reason for concern, such as values for the internal consistency of measure instruments that are considered insufficient (e.g., Cronbach's $\alpha <$ .60). Regarding interrater reliability, only one third of the studies in the review reported respective information. Thus, many of the studies that investigate EBP and/or EST education in social work that have been published so far can be considered problematic regarding the reliability of their measures.

Regarding the studies' *validity*, only 20% of the quantitative studies used validated instruments for data collection. None of the studies measured learners' skills related to EBP and/or the use of ESTs in real world settings, which is problematic considering that

both are skills that are to be applied *in practice*. A particular problem is related to the studies' disproportionately high reliance on *self-report data*. 93% of the studies based their measures solely (52%) or partly (41%) on self-report data. This high prevalence of self-report data is in line with Yaffe's (2013) assumption that only very few studies in research on social work education rely on performance data, which is problematic because individuals tend to overestimate themselves (Kruger & Dunning, 1999; Snibsøer et al., 2018). A respective example is the study by Smith, Cohen-Callow, Harnek-Hall, and Hayward (2007) which was the only study included in the systematic review that measured a construct (leaners' procedural knowledge of EBP) with self-report data as well as with performance data. They found that social work students' self-assessment of their knowledge was rather high, however, this was not resembled by a more objective measurement that captured their performance on a multiple-choice test about the methodological quality of several research scenarios. This result is in line with recent findings from Snibsøer et al. (2018), who found a low agreement between self-reported and objectively measured EBP knowledge among Canadian ($n = 154$) and Norwegian ($n = 336$) students from various healthcare disciplines. This raises doubts about the validity of many studies from the current body of research on Evidence-based Practice education in social work. Furthermore, it can be speculated that social work students tend to overestimate their EBP-related knowledge.

As for the studies' *risk of bias*, 15 (55.5%) out of 27 scored either 0, 1 or 2 (range, 0 – 4 with 0 indicating a very high risk of bias and 4 indicating a very low risk of bias) on the applied risk assessment scale (Pluye et al., 2011). Of course, risk of bias is to some extent related to the studies' reliability and validity, however, further aspects were considered in this measure (e.g., does the collected data allow to adequately address the research question). For example, most of the qualitative studies missed to appropriately

explain how their findings relate to the researchers' influence during their interactions with the participants. None of the mixed methods studies provided rationale for the necessity of conducting a mixed methods study to properly address their research question(s), let alone for the mindful integration of their qualitative and quantitative data. Furthermore, some quantitative studies had very small sample sizes. For example, one of the few studies with a controlled design had a control group with 12 and an experimental group with three participants (Webber, Currin, Groves, Hay, & Fernando, 2010). Also, other quantitative studies without controlled designs applied inferential statistics with very small sample sizes that were sometimes less than ten participants (e.g., Daniels, Torres, & Reeser, 2016; Dauenhauer, Glose, & Watt, 2015). Nevertheless, statistically significant results are reported by such studies. According to Button et al. (2013), this is problematic for at least three reasons. First, the lower the power of a study, the higher the probability that a statistically significant result does not reflect a true effect, due to a small positive predictive value. Second, it is likely that such studies report inflated effect sizes. Consequential biases will negatively affect both, the reproducibility of such studies as well as the validity of their accumulated findings (Button et al., 2013). Third, conducting such vastly underpowered studies and analyzing their data with inferential statistics can in general be considered questionable from an ethical perspective. Such studies are more likely to introduce bias into the respective body of research than to contribute to it with reliable and valid findings. It is debatable whether tax money is put to good use with such studies, considering that research is often publicly funded. Due to these findings, one may speculate that a practice might be prevalent in quantitative research on education of Evidence-based Practice in social work that Gigerenzer (2018) called the *null ritual*. The null ritual refers to the mechanical application of inferential statistics on the basis of (1) a no mean difference null

hypothesis, (2) an unspecified alternative hypothesis and (3) the acceptance of a statistically significant result on the basis of a 5% alpha level, without considering further aspects such as power. Gigerenzer (2018) argues that this entails the elimination of researchers' judgement.

Taken together, currently only very few studies provide strong evidence about the effectiveness of educational interventions to teach EBP and/or the use of ESTs in social work.

Next, the results of the second empirical study (chapter 7) are discussed. The objective of this study was to investigate effects of worked examples and an external script on social work students' fallacy recognition skills in a case-based digital learning environment.

Overall, the study found significant learning effects from the pre- to the post-test for both students' fallacy recognition skills – fallacy identification[11] and fallacy explanation – regardless of prior knowledge, time-on-task or experimental conditions. Furthermore, guided CBR (i.e., with provided information on arguments, fallacies and criteria for good argumentation; Damer, 2011) yielded significant (fallacy identification) and marginally significant (fallacy explanation) effects compared to unguided CBR (i.e., without information on arguments, fallacies and criteria for good argumentation). Effect sizes were rather small. However, students in the experimental conditions increased their fallacy recognition skills on average by approximately 1 score (range 0 – 15). Two

[11]In the second empirical study, the term *fallacy identification* was replaced with *fallacy assessment*, since that term provided a closer and more accurate match to the coding scheme that was used to code the data with regards to the skill *fallacy identification*. For reasons of coherence in relation to Part I, the term *fallacy identification* will be used in Part III instead of *fallacy assessment*.

aspects must be taken into account in that regard. First, fallacy recognition skills can be considered rather complex, since they entail numerous epistemic activities such as evidence generation, evidence evaluation and drawing conclusions (Fischer et al., 2014). Second, the intervention was relatively short (on average about 83 minutes including pre- and post-test) and could be delivered with little resources. Taking these aspects into account, results might be regarded as practically relevant (Ferguson, 2009).

Taken together, it is reasonable to consider the application of CBR in conjunction with Damer's (2011) theory *Attacking Faulty Reasoning* in the context of a digital learning environment as a promising means to facilitate fallacy recognition skills of rather advanced social work students. Of course, this assumption must be treated with caution since there was no control group for CBR.

The effects of the scaffolds, namely the worked examples and the external script, differed between (1) the process level at time 2 (during the intervention) and (2) the outcome level at time 3 (after the intervention in the post test).

At time 2 (during the intervention), both conditions, worked examples as well as the external script, yielded slightly better results for students' *fallacy identification* and *fallacy explanation skills*, compared to the control condition.

However, the main effects of the scaffolds are not interpretable since their *combination* yielded statistically significant disordinal interaction effects for students' fallacy identification as well as fallacy explanation skills. This indicates that the script as well as the worked example are effective when they are presented in isolation, but that their support decreases when they are combined. Consequentially, this means that the scaffolds are not effective in general but depend on the level of each other.

Two explanations are possible for the decrease of their effectiveness when they are combined, one refers to the (1) redundancy effect and another one more broadly to the (2) balance of extraneous and germane load with respect to learners' prior knowledge.

Combining two or more scaffolds to produce a synergistic scaffolding effect (i.e., a learning outcome that is superior compared to the learning outcomes that would be yielded when the respective scaffolds would be provided in isolation) is in general not an easy task, since the more scaffolds a learning environment entails, the more cognitively demanding it becomes (Kollar et al., 2014; Tabak, 2004). Combining the presumably already quite demanding external script (as indicated by the significantly increased time-on-task) with worked examples might have produced a split-attention condition (Sweller et al., 2011). The worked examples and the script were designed in such a way that both were comprehendible in isolation, hence, a split-attention effect would be rather unlikely, especially since the combined version could be considered rather integrated (see Spensberger et al., 2019 for pictures). However, the worked examples were erroneous examples of fallacies with regards to different criteria for good argumentation. Therefore, they implicated the related principles of the five criteria for good argumentation. The main purpose of the script was to induce and guide the application of the very same principles. Thus, combining the two scaffolds might have yielded some overlap regarding information that concerns the principles of criteria for good argumentation. A *redundancy effect* might therefore be an explanation for the negative interaction effect.

A second explanation refers to the *balance of extraneous and germane load with respect to learners' prior knowledge*. Attempts to induce germane load with additional scaffolds that are, for example, intended to induce self-explanation of underlying concepts of worked examples are sensitive to learners' prior knowledge that is necessary

for such explanations. The script did not actively prompt learners to self-explain underlying principles of the worked examples to themselves. However, it could be speculated that providing the scripts in combination with the worked examples might have induced a similar process to some extent (i.e., learners might have compared the components of the script and the respective worked examples in order to investigate and understand underlying principles). But, attempts to provide self-explanations about underlying principles might induce extraneous load rather than germane load for learners who lack (prior) knowledge that is necessary for such explanations (Paas & van Gog, 2006). This casts further doubts on the initial assumption regarding sixth semester social work students' prior knowledge about argument evaluation in social work decision making contexts.

At time 3 (post-test), no positive effects of the scaffolds were observed. Quite the contrary, a negative main effect of the script was observed for fallacy identification as well as for fallacy explanation, indicating that the script had a negative effect on students' fallacy recognition skills compared to no intervention. In general, it is possible that learners need more time to discard their older suboptimal learning strategies in favor of the newly acquired ones in order to benefit from them. Often, learners return to their older learning strategies immediately after the training of new ones (Renkl, 2008). Nevertheless, the results are contrary to much empirical evidence that suggests that worked examples as well as scripts can have substantial positive effects on students' learning outcomes (e.g., Atkinson, Renkl, & Merrill, 2003; Sweller, Ayres, & Kalyuga, 2011; Vogel, Wecker, Kollar, & Fischer, 2017). Thus, it is reasonable to assume that the scaffolds were not optimally designed. The following closer examination of the scaffolds supports this assumption.

Regarding the *worked examples*, especially three aspects might have been problematic: (1) The fact that they did not show the solution process in detail, (2) the lack of inducement of germane load and (3) the fact that they were erroneous.

First, the applied worked examples did not show the *solution processes* that have to be undertaken in order to identify and explain a fallacy. They were merely examples of similar fallacies. It is thus possible that they were not rich enough, thereby still leaving learners with the necessity to search the problem space for aspects that were relevant for schema acquisition (Renkl, 2014). That way, the worked examples might not have contributed sufficiently regarding the release of working memory capacity for an increase of germane load. It is an interesting yet open question how to model worked examples, so that they properly resemble the solution path to identify and explain fallacies in a social work decision making context.

Second, no additional support that was tailored towards the increase of germane load was provided while working with worked examples. The worked examples might have led to a decreased amount of extraneous load in learners' working memories (compared to no worked examples, which might have induced the need for the learner to search the problem space for aspects that are relevant for schema acquisition), thereby leaving enough capacity for an adequate amount of germane load. However, the remaining free capacity does not automatically lead to an increased amount of germane load (de Jong, 2010). Hence, additional support tailored towards the increase of germane load might have been necessary. On the one hand, worked examples are a scaffold that allows a learner to remain rather passive (Renkl, 2014). On the other hand, fallacy recognition skills (and likewise their respective principles) can be considered rather complex. Therefore, the application of additional support such as self-explanation prompts (e.g., Atkinson et al., 2003; Hilbert, Renkl, Kessler, & Reiss, 2008; Schworm &

Renkl, 2007) or feedback (e.g., Kopp, Stark, & Fischer, 2008; Kopp, Stark, Kühne-Eversmann, & Fischer, 2009) might be beneficial in helping learners to deeply process and understand the underlying principles of the worked examples that are supposed to guide learners in the application of fallacy recognition skills in social work decision making contexts. For example, research by Schworm and Renkl (2007) suggests that when the goal is to foster skills, worked examples might be more effective when they are combined with self-explanation prompts. Therefore, such support methods might be promising candidates regarding the intentional increase of germane load while working with worked examples.

Third, the worked examples can be considered *erroneous*, since they provided learners with an argument that violated criteria for good argumentation in order to show the learner how a fallacy that violates a specific criterion for good argumentation (Damer, 2011) would look like. This is not problematic per se, however, empirical evidence from Große and Renkl (2007) in the context of a well-structured domain (solving probability problems) suggests that erroneous worked examples are not favorable for learners with little prior knowledge. Therefore, it is possible that students' prior knowledge on argument evaluation (e.g., due to field practice; see chapter 5.2) was overestimated with regards to the design of the worked examples and that the worked examples were not ideally tailored towards learners' prerequisites. In other words, *correct* worked examples might have been more beneficial considering learners' prior knowledge. This assumption is in line with learners overall mediocre scores regarding fallacy identification and their generally low scores with respect to fallacy explanation. Furthermore, Große and Renkl (2007) found that erroneous worked examples can have detrimental effects on learners' principle-based self-explanations. Hence, it might have been difficult for learners to

appropriately explain the underlying principles of the worked examples to themselves, which is important for schema acquisition (Renkl, 2017).

As for the results of the *external script*, four aspects are especially interesting: (1) the optimal external scripting level principle (one of the principles proposed in the Script Theory of Guidance, see chapter 4.3.2.2), (2) learners' significantly higher time-on-task, the possibility of (3) overscripting and the script's (4) level of coercion.

First, the script might not have been in line with the *optimal external scripting level principle* (Fischer, Kollar, Stegmann, & Wecker, 2013), meaning that it did not provide the learners with an optimal Zone of Proximal Development (Vygotsky, 1978). As stated before, it might be possible that students' prior knowledge was overestimated. This assumption seems especially reasonable in consideration of the aforementioned complexity of fallacy recognition skills. Future attempts in designing scripts to facilitate social work students' fallacy recognition skills should therefore take students' related internal scripts into account more thoroughly, for example by analyzing them upfront (Kollar, Fischer, & Slotta, 2007).

Second, students in script conditions needed significantly more *time-on-task* compared to the other conditions. Therefore, a possible explanation for the disappointing results of the script could be that students needed a considerable amount of time to understand the script (i.e., how it functions and how they should use it). This assumption is supported by the fact that students were not explicitly trained in using the script. The learning environment provided a description of how to use the script, however, this might not have been enough to convey students in how to use it appropriately. This assumption would be in line with the presumption that students' prior knowledge might have been overestimated, which might have led to further insecurities of the students regarding the

question what the script was supposed to support. In this regard, Dillenbourg (2002) suggests that experimental studies on the effectiveness of scripts require learners to be trained in how to use them.

A third possible explanation that is to some extent related to the increased time-on-task in the script conditions is that the script provided too much information that had to be cognitively processed by the learners. Hence, the learning effects of the script might have been hampered due to *overscripting* (Dillenbourg, 2002). In other words, processing all the information that was provided by the script might have overstrained the learner's working memory by putting too much extraneous load on it, thereby leaving little capacity for germane load which is necessary for schema-acquisition. The script consisted of an instructional introduction (e.g., provide your explanation in the textbox if the claim violates a criterion for good argumentation), five headers, each of which referring to one of the criteria for good argumentation (e.g., structure), respective prompts (e.g., "claim X violates the *structure* criterion because …") and a textbox (see Spensberger, Kollar, & Pankofer, 2019 for pictures). Thus, especially the prompts could be considered as too much information, since they were provided for each criterion, even though not all claims violated all criteria (which, in addition, could be considered as potentially misleading).

A fourth possible explanation which is partially related to overscripting refers to the high level of *coercion* of the script. The script applied in the second study prompted the students to perform cognitive operations in a specific order and learners had to strictly follow the sequence of the prompts, regardless of whether a criterion of a prompt was relevant for a given claim or not. For example, a learner might have evaluated a claim against the structure criterion (the first of the five criteria) in such a way that she decided that the claim violated the very criterion. Afterwards, she would still have to evaluate the

claim against the remaining four other criteria. This might have been too coercive, thereby leaving the students with little degrees of freedom in following the script, which might have impaired leaners' self-regulation (Wecker, Kollar, Fischer, & Prechtl, 2010) and/or led to a possible decrease of students' motivation (Dillenbourg, 2002). One possible solution for this problem could be the application of more adaptable scripts (Noroozi, Kirschner, Biemans, & Mulder, 2018; Wang, Kollar, & Stegmann, 2017), that allow the learner to adjust them to her needs. Such a script could, for example, let the learner directly pick the needed criterion from a drop-down menu, or hide criteria two to five and ask the learner, after evaluating criterion one, if criterion two (and respectively three, four and five) is needed and should therefore be shown.

9. Implications for Social Work Research and Practice

This chapter outlines potential implications for social work research and practice derived from the results of the two empirical studies (chapter 6 and 7) and the subsequent general discussion of their results (chapter 8).

9.1 Implications for Social Work Research

The first study (chapter 6) was among the first to provide a systematic overview of research on education of Evidence-based Practice in social work.

The findings of the first study suggest, that guided student-centered approaches are promising in facilitating Evidence-based Practice in social work. However, this is also the case for other approaches such as teacher-centered ones. It is of course possible, that "anything goes" in that regard. However, this would contrast much empirical evidence from educational psychology which suggests that teacher-centered approaches tend to be more effective for learners with little prior knowledge, while student-centered

approaches are likely to be more effective for more advanced learners (Kollar & Fischer, 2019). The fact that the review could not provide evidence regarding a general superiority of student-centered approaches is especially concerning considering the strong push in the conceptual literature of disciplinary social work to apply Problem-based Learning in EBP-education under any circumstances (e.g., Drake et al., 2007; Drisko & Grady, 2018; Franklin, 2007; Gambrill, 1999, 2007). Hence, further studies on education of Evidence-based Practice in social work are needed. They should compare different instructional approaches regarding their effectiveness on learning outcomes under different circumstances (e.g., learners with little prior knowledge vs. advanced learners).

Furthermore, the findings suggest that the current body of research on education of Evidence-based Practice in social work does not provide strong evidence to support causal inferences regarding the effectiveness of instructional approaches, due to the scarce amount of controlled designs. Moreover, the findings cast considerable doubts on the validity and reliability of many applied measure instruments in research on education of Evidence-based Practice in social work. A specific threat to the validity of the respective current body of research is the prevalence of self-report data. Therefore, there is a strong need for experimental research on education of Evidence-based Practice in social work. Such future studies need to apply valid and reliable measure instruments that focus on performance data, preferably in real world settings. Furthermore, data analyses of quantitative studies should be aware of questionable practices such as the null ritual (see chapter 8.2; Gigerenzer, 2018) and consider the mindful application of inferential statistics.

The second study (chapter 7) was among the first to empirically investigate the facilitation of fallacy recognition skills in a decision-making context in general and particularly among social work students. This study has several implications for social work research from a methodological as well as from a theoretical perspective, despite its rather disappointing results.

The results of the first empirical study showed that experimental designs are hardly applied in research related to education of Evidence-based Practice in social work, especially regarding the comparison of different instructional methods. The second study provides an example of such a methodological approach. As outlined earlier in this doctoral thesis, it is reasonable to assume that research on social work education might benefit from such methodologies. This assumption is supported by arguments and evidence on the importance of experimental research for the development of educational practice (e.g., Bruer, 1994; McGilly, 1998).

Furthermore, the second empirical study introduced several theoretical concepts from research in educational psychology (e.g., scripts) to research on education in social work. Respective scaffolds can yield significant effects on learners' fallacy recognition skills at the process level when they are presented in isolation, as indicated by the results of the second empirical study. Results at the outcome level were rather disappointing, however, much empirical evidence suggests that such scaffolds can yield effective learning outcomes (e.g., Atkinson et al., 2003; Sweller et al., 2011; Vogel et al., 2017). It is thus likely that the scaffolds that were applied in the second study were not effective due to a suboptimal design. Further studies are needed to investigate how these scaffolds must be designed in order to be effective in education of Evidence-based Practice in social work. In this regard, several suggestions for future studies can be drawn from the explanations that were provided in the general discussion (chapter 8.2).

As for *worked examples*, three possible explanations regarding their disappointing results were provided, namely the fact that they did not outline the solution path, the lack of support to induce germane load and doubts regarding the appropriateness of erroneous worked examples concerning learners' prior knowledge. It is thus reasonable to assume that correct worked examples that explicitly show the solution path and that are paired with additional support such as prompts that are designed to enhance self-explanation processes might be more beneficial for the facilitation of social work students' fallacy recognition skills (e.g., Atkinson et al., 2003; Hilbert et al., 2008; Schworm & Renkl, 2007). Other options are heuristic worked examples that provide underlying principles (Renkl et al., 2009) or, when erroneous worked examples are applied, to combine the worked examples with elaborated feedback (Kopp et al., 2009).

Likewise, four explanations were provided regarding the effects of the *external script*. It was argued that they might not have been optimally designed with respect to students' prerequisites (i.e., their internal scripts), that the design of the script in combination with lack of training made it difficult for the learners to understand how the script worked (i.e., trying to understand the script might have put heavy demands on the learners' working memory) as indicated by the significantly higher time-on-task in the script conditions, that the amount of information provided by the script might have led to overscripting and that their high degree of coercion might have induced a decrease in learners' motivation and/or impaired their self-regulation. Therefore, future research should align external scripts more properly to students' prerequisites by analyzing their internal scripts upfront (e.g., Kollar et al., 2007). Training on how to use the script should be provided prior to the experiment (Dillenbourg, 2002) and adaptable scripts (Noroozi et al., 2018; Wang et al., 2017) with higher degrees of freedom might be preferable

compared to rather static and strict scripts with lower degrees of freedom. Furthermore, controlling for learners' motivation should be considered, especially when applied scripts are rather coercive.

Furthermore, and with regard to both scaffolds, it is important to consider the complexity of fallacy recognition skills on the one hand and the rather short interventions of the second empirical study on the other hand. It is possible that the provided interventions were too short to yield significant learning outcomes and that longer interventions (e.g., over the course of a whole semester) might yield greater effects in that regard.

Combining scaffolds to achieve *synergistic scaffolding* effects is a particularly demanding task (Tabak, 2004). It was argued that the combination of the worked examples and the script might have led to a redundancy effect (at least to some extent). A second rather speculative explanation was that the combination of the scaffolds might have induced cognitive processes that were not intended (e.g., attempts to use the script to infer underlying principles of the worked examples), thereby inducing extraneous instead of germane load due to an underestimation of learners' prior knowledge (de Jong, 2010). Consequently, combined scaffolds must not feature an overlap with regards to their content to avoid redundancy effects.

9.2 Implications for Social Work Practice

It is difficult to derive recommendations for social work practice from both empirical studies, given their rather inconclusive results with regards to questions of effectiveness of instructional approaches and scaffolds.

The findings of the *first empirical study* (chapter 6) suggest that student- and teacher-centered approaches might be equally effective. However, this is questionable

with respect to the validity and reliability of the current body of research on education of Evidence-based Practice in social work, as well as with respect to empirical evidence from fields other than social work (Kollar & Fischer, 2019). Hence, social work educators who intend to facilitate skills regarding the process of EBP and/or the use of EST/s should critically reflect the current social work literature and draw on research from fields other than social work such as various healthcare professions (see for example, Aglen, 2016; Dizon, Grimmer-Somers, & Kumar, 2012; Horntvedt, Nordsteien, Fermann, & Severinsson, 2018; Kyriakoulis et al., 2016; Larsen, Terkelsen, Carlsen, & Kristensen, 2019; Patelarou et al., 2017 for recent reviews).

In the *second empirical study* (chapter 7), students demonstrated a moderate performance on fallacy identification and a rather weak performance on fallacy explanation on average. Hence, the second study provides empirical evidence for the need to educate social work students in fallacy recognition skills, as resembled by several respective calls in the conceptual social work literature (e.g., Gambrill, 2014; Gambrill & Gibbs, 2017). These results are in line with the currently moderate implementation status of EBP (e.g., Knight, 2013, 2015), considering the fact that many scholars argue that critical thinking skills such as fallacy recognition skills are a prerequisite for EBP (e.g., Aglen, 2016; Gambrill, 2019; Profetto-McGrath, 2005). Therefore, it could be speculated that the results of the second empirical study support the (theory-driven) assumption that such skills are an important prerequisite for EBP. It is of course not to say that an increase of social work students' fallacy recognition skills will automatically lead to a wider implementation of EBP in social work practice. However, they might indeed be regarded as one aspect that could contribute to the further implementation of EBP. Also, the general importance of argument evaluation for social work decision making contexts

(see chapters 3.1 and 3.2) clearly suggest the need to promote the facilitation of fallacy recognition skills and to conduct further research on how to best teach such skills in social work education. Towards that end, Damer's (2011) theory *Attacking Faulty Reasoning* is a promising theoretical account, as indicated by the results of the second study. It can furthermore be speculated that CBR is a promising instructional approach in that regard.

10. Conclusion

The goals of this doctoral thesis are to provide an overview of research on Evidence-based Practice in social work, to assess the respective studies with regards to the question how to teach its two main approaches in social work – namely the process of Evidence-based Practice and Empirically Support Treatments –and to investigate the facilitation of social work students' fallacy recognition skills, which are seen as a prerequisite for the process of Evidence-based Practice.

Unfortunately, the answers to the first two questions (the first empirical study; chapter 6) could not provide extensive recommendations for social work practitioners (i.e., educators). However, they suggest that the currently available evidence for education of Evidence-based Practice in social work is rather thin, thereby underlining the importance for disciplinary social work to conduct further research on Evidence-based Practice education. Towards that end, a necessary next step for social work research is to apply controlled designs with the goal to compare the effectiveness of different instructional approaches while taking learners' prerequisites into account.

Answering the third question (second empirical study; chapter 7) yielded – again – results that did not allow for rich recommendations for social work practice. However, the study provided empirical evidence for a request to social work education that so far has been merely based on reasonable assumptions (i.e., promoting the facilitation of fallacy recognition skills in social work). It provided a first step regarding the empirical investigation of how to best teach social work students' fallacy recognition skills in decision making contexts. Empirical evidence in this regard is still scarce and it is an open question for social work research how to design scaffolds in such a way, that they can

enhance students' learning outcomes on their fallacy recognition skills beyond the use of Case-based Reasoning as an instructional approach.

Taken together, the ongoing call for Evidence-based Practice in social work implies the need for Evidence-based Education (i.e., *use* best currently available research evidence and *establish* sound evidence when evidence is lacking or when it is of weak nature; Davies, 1999; Slavin, 2008) regarding the education of Evidence-based Practice approaches as well as related prerequisites (such as fallacy recognition skills), in order to promote the implementation of Evidence-based Practice in professional social work (see chapter 2). This is of utmost importance, since the respective evidence-base can currently be considered rather thin.

Hopefully, this doctoral thesis will contribute to the future increase of this evidence-base by providing empirical evidence that underlines the need for such an expansion, and by introducing promising methodologies and theoretical concepts as well as ideas derived from empirical evidence for further related research. A resultant enhanced evidence-base "might enable social work educators to practice what they preach in the form of evidence-based social work education" (Yaffe, 2013, p. 527).

References

References marked with an asterisk indicate studies included in the systematic review.

Aarons, G. A. (2004). Mental health provider attitudes toward adoption of evidence-based practice: The Evidence-Based Practice Attitude Scale (EBPAS). *Mental Health Services Research*, *6*, 61–74.

Aarons, G. A., Glisson, C., Hoagwood, K., Kelleher, K., Landsverk, J., & Cafri, G. (2010). Psychometric properties and U.S. national norms of the Evidence-Based Practice Attitude Scale (EBPAS). *Psychological Assessment*, *22*, 356–365. https://doi.org/10.1037/a0019188

Ablon, J. S., & Jones, E. E. (2002). Validity of controlled clinical trials of psychotherapy: Findings from the NIMH treatment of depression collaborative research program. *The American Journal of Psychiatry*, *159*(5), 775–783. https://doi.org/10.1176/appi.ajp.159.5.775

Adams, K. B., Matto, H. C., & LeCroy, C. W. (2009). Limitations of evidence-based practice for social work education: Unpacking the complexity. *Journal of Social Work Education*, *45*(2), 165–186. https://doi.org/10.5175/JSWE.2009.200700105

*Ager, R., Roahen-Harrison, S., Toriello, M. P. H., Morse, P., Morse, E., Carney, L., . . . Kissinger, P. (2005). A study on the effectiveness of a brief motivational enhancement therapy training. In E. Edmundson & D. McCarty (Eds.), *Implementing evidence-based practices for treatment of alcohol and drug disorders* (pp. 15–32). Binghampton: Haworth Medical Press. https://doi.org/10.1300/J069v24S01_02

Aglen, B. (2016). Pedagogical strategies to teach bachelor students evidence-based practice: A systematic review. *Nurse Education Today*, *36*, 255–263. https://doi.org/10.1016/j.nedt.2015.08.025

Albus, S., & Micheel, H.-G. (2012). Entmündigung der Praxis? Wirkungsorientierung und Evidence Based Practice. In K. Unzicker & G. Hessler (Eds.), *Öffentliche Sozialforschung und Verantwortung für die Praxis: Zum Verhältnis von Sozialforschung, Praxis und Öffentlichkeit* (Vol. 1, pp. 179–197). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-531-94221-6_9

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist*, *51*(4), 355–365. https://doi.org/10.1037/0003-066X.51.4.355

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *Psychology of learning and motivation* (Vol. 2, pp. 89–195). Academic Press. https://doi.org/10.1016/S0079-7421(08)60422-3

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research*, *70*(2), 181–214. https://doi.org/10.3102/00346543070002181

Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology*, *95*(4), 774–783. https://doi.org/10.1037/0022-0663.95.4.774

Baddeley, A. D., & Hitch, G. (1968). Working memory. In K. W. Spence & J. T. Spence (Eds.), *Psychology of learning and motivation* (Vol. 8, pp. 47–89). Academic Press. https://doi.org/10.1016/S0079-7421(08)60452-1

Barlow, D. H. (2010). Negative effects from psychological treatments: A perspective. *The American Psychologist*, *65*(1), 13–20. https://doi.org/10.1037/a0015643

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*, *128*(4), 612–637. https://doi.org/10.1037/0033-2909.128.4.612

Barrows, H. S., & Tamblyn, R. M. (1980). *Problem-based learning: An approach to medical education*. *Springer series on medical education: v. 1*. New York: Springer Pub. Co.

Bates, M. (2006). A critically reflective approach to evidence-based practice: A sample of school social workers. *Canadian Social Work Review*, *23*(1/2), 95–109.

Bauhoff, V., Huff, M., & Schwan, S. (2012). Distance matters: Spatial contiguity effects as trade-off between gaze switches and memory load. *Applied Cognitive Psychology*, *26*(6), 863–871. https://doi.org/10.1002/acp.2887

Bellamy, J. L., Bledsoe, S. E., & Traube, D. E. (2006). The current state of evidence-based practice in social work: A review of the literature and qualitative analysis of expert interviews. *Journal of Evidence-Informed Social Work*, *3*(1), 23–48. https://doi.org/10.1300/J394v03n01_02

*Bellamy, J. L., Fang, L., Bledsoe, S. E., Manuel, J. I., Dinata, E., & Mullen, E. J. (2013). Agency-university partnership for evidence-based practice in social work. *Journal of Evidence-Informed Social Work*, *10*(2), 73–90. https://doi.org/10.1080/15433714.2011.581545

*Bender, K., Altschul, I., Yoder, J., Parrish, D., & Nickels, S. J. (2014). Training social work graduate students in the evidence-based practice process. *Research on Social Work Practice*, *24*(3), 339–348. https://doi.org/10.1177/1049731513506614

Blayney, P., Kalyuga, S., & Sweller, J. (2015). Using cognitive load theory to tailor instruction to levels of accounting students' expertise. *Journal of Educational Technology & Society*, *18*(4), 199–210.

*Bledsoe-Mansori, S. E., Manuel, J. I., Bellamy, J. L., Fang, L., Dinata, E., & Mullen, E. J. (2013). Implementing evidence-based practice: Practitioner assessment of an agency-based training program. *Journal of Evidence-Informed Social Work*, *10*(2), 73–90. https://doi.org/10.1080/15433714.2011.581545

Bohart, A. C., O'Hara, M., & Leitner, L. M. (1998). Empirically violated treatments: Disenfranchisement of humanistic and other psychotherapies. *Psychotherapy Research*, *8*(2), 141–157. https://doi.org/10.1093/ptr/8.2.141

Booysen, F., Mbecke, P., Gouveia, A. de, Manomano, T., & Tanga, P. (2019). Attitudes to evidence-based practice among social work practitioners in south africa. *Southern African Journal of Social Work and Social Development*, *31*(1). https://doi.org/10.25159/2415-5829/4160

Borrmann, S., & Thiessen, B. (2016). *Wirkungen Sozialer Arbeit: Potentiale und Grenzen der Evidenzbasierung für Profession und Disziplin* (1st ed.). *Theorie, Forschung und Praxis der Sozialen Arbeit: v.12*. Leverkusen-Opladen: Budrich Barbara.

Bruer, J. T. (1994). *Schools for thought: A science of learning in the classroom* (1. MIT press paperback ed.). *A Bradford book*. Cambridge,Mass.: M I T Press.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Carmien, S., Kollar, I., Fischer, G., & Fischer, F. (2007). The interplay of internal and external scripts - A distributed cognition perspective. In F. Fischer, I. Kollar, H. Mandl, & J. M. Haake (Eds.), *Scripting computer-supported collaborative learning: Cognitive, computational and educational perspectives* (pp. 303–326). New York: Springer.

Carroll, W. M. (1994). Using worked examples as an instructional support in the algebra classroom. *Journal of Educational Psychology*, *86*(3), 360–367. https://doi.org/10.1037/0022-0663.86.3.360

Cartwright, N., & Munro, E. (2010). The limitations of randomized controlled trials in predicting effectiveness. *Journal of evaluation in clinical practice*, *16*(2), 260–266. https://doi.org/10.1111/j.1365-2753.2010.01382.x

Castelnuovo, G. (2010). Empirically supported treatments in psychotherapy: Towards an evidence-based or evidence-biased psychology in clinical settings? *Frontiers in Psychology*, *1*, 27. https://doi.org/10.3389/fpsyg.2010.00027

Castelnuovo, G., Faccio, E., Molinari, E., Nardone, G., & Salvini, A. (2004). A critical review of Empirically Supported Treatments (ESTs) and common factors perspective in psychotherapy. *Brief Strategic and Systemic Therapy European Review*, *1*, 208–224.

Cawood, N. D. (2010). Barriers to the use of evidence-supported programs to address school violence. *Children & Schools*, *32*(3), 143–149. https://doi.org/10.1093/cs/32.3.143

Cerpa, N., Chandler, P., & Sweller, J. (1996). Some conditions under which integrated computer-based training software can facilitate learning. *Journal of Educational Computing Research*, *15*(4), 345–367. https://doi.org/10.2190/MG7X-4J8N-CKYR-P06T

Chambless, D. L. (1996). In defense of dissemination of empirically supported psychological interventions. *Clinical Psychology: Science and Practice*, *3*(3), 230–235. https://doi.org/10.1111/j.1468-2850.1996.tb00074.x

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, *66*(1), 7–18. https://doi.org/10.1037/0022-006X.66.1.7

Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology*, *52*, 685–716. https://doi.org/10.1146/annurev.psych.52.1.685

Chambless, D. L., Baker, M., Baucom, D. H., Beutler, L., Calhoun, K. S., Crits-Christoph, P., . . . Woody, S. R. (1998). Update on empirically validated therapies II. *The Clinical Psychologist*, *51*(1), 3–16.

Chambless, D. L., Sanderson, W. C., Shoham, V., Bennett Johnson, S., Pope, K. S., Crits-Christoph, P., . . . McCurry, S. (1996). An update on empirically validated therapies. *The Clinical Psychologist*, *49*(2), 5–18.

Chandler, P., & Sweller, J. (1996). Cognitive load while learning to use a computer program. *Applied Cognitive Psychology*, *10*(2), 151–170. https://doi.org/10.1002/(SICI)1099-0720(199604)10:2<151::AID-ACP380>3.0.CO;2-U

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55–81. https://doi.org/10.1016/0010-0285(73)90004-2

Chi, M. T.H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive science*, *13*(2), 145–182. https://doi.org/10.1207/s15516709cog1302_1

Claridge, J. A., & Fabian, T. C. (2005). History and development of evidence-based medicine. *World Journal of Surgery*, *29*(5), 547–553. https://doi.org/10.1007/s00268-005-7910-1

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hoboken: Taylor and Francis.

Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology*, *79*(4), 347–362. https://doi.org/10.1037/0022-0663.79.4.347

Copi, I. M., & Burgess-Jackson, K. (1996). *Informal logic* (3. ed.). Upper Saddle River, N.J.: Prentice Hall.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *The Behavioral and Brain Sciences*, *24*(1), 87-114; discussion 114-85. https://doi.org/10.1017/s0140525x01003922

Damer, T. E. (2011). *Attacking faulty reasoning: A practical guide to fallacy-free arguments* (7. ed.). Boston, MA: Wadsworth Cengage Learning.

*Daniels, R. A., Torres, D., & Reeser, C. (2016). Where words fail, music speaks: A mixed method study of an evidence-based music protocol. *Journal of Evidence-Informed Social Work*, *13*(6), 535–551. https://doi.org/10.1080/23761407.2015.1111825

*Dauenhauer, J. A., Glose, S., & Watt, C. (2015). Design, delivery, and outcomes from an interprofessional fall prevention course. *Gerontology & Geriatrics Education*, *36*(3), 278–301. https://doi.org/10.1080/02701960.2015.1031891

Davies, P. (1999). What is Evidence-based education? *British Journal of Educational Studies*, *47*(2), 108–121. https://doi.org/10.1111/1467-8527.00106

De Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science*, *38*(2), 105–134. https://doi.org/10.1007/s11251-009-9110-0

De Jong, T., & Ferguson-Hessler, M. G. M. (1996). Types and qualities of knowledge. *Educational Psychologist*, *31*(2), 105–113.

Deutscher Berufsverband für Soziale Arbeit e.V. (2014). Berufsethik des DBSH. Retrieved from https://www.dbsh.de/profession/berufsethik/berufsethik-des-dbsh.html

Dickersin, K., Straus, S. E., & Bero, L. A. (2007). Evidence based medicine: Increasing, not dictating, choice. *BMJ (Clinical Research Ed.)*, *334 Suppl 1*, s10. https://doi.org/10.1136/bmj.39062.639444.94

Dillenbourg, P. (2002). Over-scripting CSCL. In P. A. Kirschner (Ed.), *Three worlds of CSCL: Can we support CSCL* (pp. 61–91). Heerlen: University of the Netherlands.

Dizon, J. M. R., Grimmer-Somers, K. A., & Kumar, S. (2012). Current evidence on evidence-based practice training in allied health: A systematic review of the literature. *International Journal of Evidence-Based Healthcare*, *10*(4), 347–360. https://doi.org/10.1111/j.1744-1609.2012.00295.x

Dochy, F., Segers, M., van den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: A meta-analysis. *Learning and Instruction*, *13*(5), 533–568. https://doi.org/10.1016/S0959-4752(02)00025-7

Drake, B., Hovmand, P., Jonson-Reid, M., & Zayas, L. H. (2007). Adopting and teaching evidence-based practice in master's-level social work programs. *Journal of Social Work Education*, *43*(3), 431–446. https://doi.org/10.5175/JSWE.2007.200600625

Drisko, J., & Grady, M. D. (2018). Teaching evidence-based practice using cases in social work education. *Families in Society: The Journal of Contemporary Social Services*, *99*(3), 269–282. https://doi.org/10.1177/1044389418785331

*Ducharme, P., Rober. A, & Wharff, E. (2015). Building confidence in social work interns through an evidence-based practice seminar during field education. *Field Educator*, *5*.

Ekeland, T. J., Bergem, R., & Myklebust, V. (2019). Evidence-based practice in social work: Perceptions and attitudes among Norwegian social workers. *European Journal of Social Work*, *22*(4), 611–622. https://doi.org/10.1080/13691457.2018.1441139

Emparanza, J. I., Cabello, J. B., & Burls, A. J. E. (2015). Does evidence-based practice improve patient outcomes? An analysis of a natural experiment in a spanish hospital. *Journal of Evaluation in Clinical Practice*, *21*(6), 1059–1065. https://doi.org/10.1111/jep.12460

Engelke, E., Borrmann, S., & Spatscheck, C. (2018). *Theorien der Sozialen Arbeit: Eine Einführung* (7., überarbeitete und erweiterte Auflage). Freiburg im Breisgau: Lambertus.

Evidence-based Medicine Working Group (1992). Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA*, *268*(17), 2420–2425. https://doi.org/10.1001/jama.1992.03490170092032

Eysink, T. H. S., de Jong, T., Berthold, K., Kolloffel, B., Opfermann, M., & Wouters, P. (2009). Learner performance in multimedia learning arrangements: An analysis across instructional approaches. *American Educational Research Journal*, *46*(4), 1107–1149. https://doi.org/10.3102/0002831209340235

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*(5), 532–538. https://doi.org/10.1037/a0015808

Fischer, F., Chinn, C. A., Engelmann, K. F., & Osborne, J. (Eds.). (2018). *Scientific reasoning and argumentation: The roles of domain-specific and domain-general knowledge*. New York, London: Routledge.

Fischer, F., Hmelo-Silver, C. E., Goldman, S. R., & Reimann, P. (Eds.). (2018). *International handbook of the learning sciences*. New York, London: Routledge.

Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist*, *48*(1), 56–66. https://doi.org/10.1080/00461520.2012.748005

Fischer, F., Kollar, K., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., . . . Eberle, J. (2014). Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda in education. *Frontline Learning Research*. *2*(3), 28–45. https://doi.org/10.14786/flr.v2i2.96

Franklin, C. (2007). Teaching evidence-based practices: Strategies for implementation: A response to Mullen et al. and Proctor. *Research on Social Work Practice*, *17*(5), 592–602. https://doi.org/10.1177/1049731507301525

Gambrill, E. (1994). Social work research: Priorities and obstacles. *Research on Social Work Practice*. (3), 359–388.

Gambrill, E. (1999). Evidence-based clinical behavior analysis, evidence-based medicine and the Cochrane collaboration. *Journal of Behavior Therapy and Experimental Psychiatry*, *30*(1), 1–14. https://doi.org/10.1016/S0005-7916(98)00035-4

Gambrill, E. (1999). Evidence-based practice: An alternative to authority-based practice. *Families in Society: The Journal of Contemporary Social Services*, *80*(4), 341–350. https://doi.org/10.1606/1044-3894.1214

Gambrill, E. (2006). Evidence-based practice and policy: Choices ahead. *Research on Social Work Practice*, *16*(3), 338–357. https://doi.org/10.1177/1049731505284205

Gambrill, E. (2007). Transparency as the route to evidence-informed professional education. *Research on Social Work Practice*, *17*(5), 553–560. https://doi.org/10.1177/1049731507300149

Gambrill, E. (2010). Evidence-based practice and the ethics of discretion. *Journal of Social Work*, *11*(1), 26–48. https://doi.org/10.1177/1468017310381306

Gambrill, E. (2011). *Propaganda in the helping professions*. Oxford: Oxford University Press USA.

Gambrill, E. (2012). Uses of history in creating new futures: A science-informed social work. *Research on Social Work Practice*, *22*(5), 481–491. https://doi.org/10.1177/1049731512443714

Gambrill, E. (2013). *Social work practice: A critical thinker's guide* (3rd ed.). New York: Oxford University Press.

Gambrill, E. (2014). Social work education and avoidable ignorance. *Journal of Social Work Education*, *50*(3), 391–413. https://doi.org/10.1080/10437797.2014.916937

Gambrill, E. (2019). *Critical thinking and the process of evidence-based practice*. New York, NY: Oxford University Press.

Gambrill, E., & Gibbs, L. (2017). *Critical thinking for helping professionals: A skills-based workbook*. New York, NY: Oxford University Press.

Garfield, S. L. (1996). Some problems associated with "validated" forms of psychotherapy. *Clinical Psychology: Science and Practice*, *3*(3), 218–229. https://doi.org/10.1111/j.1468-2850.1996.tb00073.x

Geiser, K. (2015). *Problem- und Ressourcenanalyse in der Sozialen Arbeit: Eine Einführung in die Systemische Denkfigur* (6. korrigierte Auflage). Luzern, Freiburg i. Br.: interact Verlag; Lambertus Verlag GmbH.

Ghanem, C., Lawson, T. R., Pankofer, S., Maragkos, M., & Kollar, I. (2017). The diffusion of evidence-based practice: Reviewing the evidence-based practice networks in the United States and German-speaking countries. *Journal of Evidence-Informed Social Work*, *14*(2), 86–118. https://doi.org/10.1080/23761407.2017.1298074

Gibbs, L. (2002). *Evidence-based practice for the helping professions: A practical guide with integrated*. Pacific Grove: Brooks Cole.

Gibbs, L., & Gambrill, E. (2002). Evidence based practice: Counterarguments to objections. *Research on Social Work Practice*, *12*(3), 452–476. https://doi.org/10.1177/1049731502012003007

Gigerenzer, G. (2018). Statistical rituals: The replication delusion and how we got there. *Advances in Methods and Practices in Psychological Science*, *1*(2), 198–218. https://doi.org/10.1177/2515245918771329

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*, 451–482. https://doi.org/10.1146/annurev-psych-120709-145346

Gijbels, D., Dochy, F., van den Bossche, P., & Segers, M. (2005). Effects of problem-based learning: A meta-analysis from the angle of assessment. *Review of Educational Research*, *75*(1), 27–61. https://doi.org/10.3102/00346543075001027

Gitterman, A., & Knight, C. (2013). Evidence-guided practice: Integrating the science and art of social work. *Families in Society: The Journal of Contemporary Social Services*, *94*(2), 70–78. https://doi.org/10.1606/1044-3894.4282

Gräsel, C., & Mandl, H. (1993). Föderung des Erwerbs diagnostischer Strategien in fallbasierten Lernumgebungen. *Unterrichtswissenschaft*, *21*, 355–370.

Gray, M., Joy, E., Plath, D., & Webb, S. A. (2012). Implementing evidence-based practice. *Research on Social Work Practice*, *23*(2), 157–166. https://doi.org/10.1177/1049731512467072

Gray, M., Joy, E., Plath, D., & Webb, S. A. (2015). What supports and impedes evidence-based practice implementation? A survey of Australian social workers. *British Journal of Social Work*, *45*(2), 667–684. https://doi.org/10.1093/bjsw/bct123

Gray, M., & McDonald, C. (2006). Pursuing good practice? *Journal of Social Work*, *6*(1), 7–20. https://doi.org/10.1177/1468017306062209

Gray, M., Plath, D., & Webb, S. A. (2009). *Evidence-based social work: A critical stance* (repr). London: Routledge.

Green, J. (2000). The role of theory in evidence-based health promotion practice. *Health Education Research*, *15*(2), 125–129. https://doi.org/10.1093/her/15.2.125

*Gromoske, A. N., & Berger, L. K. (2017). Replication of a continuing education workshop in the evidence-based practice process. *Research on Social Work Practice*, *27*(6), 676–682. https://doi.org/10.1177/1049731515597477

Große, C. S., & Renkl, A. (2004). Learning from worked examples: What happens if errors are included. *Instructional design for effective and enjoyable computer-supported learning*, 356–364.

Große, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning and Instruction*, *17*(6), 612–634. https://doi.org/10.1016/j.learninstruc.2007.09.008

Gruber, H., Law, L., Mandl, H., & Renkl, A. (1996). Situated learning and transfer. In P. Reimann & H. Spada (Eds.), *Journals of related interest. Learning in humans and machines: Towards an interdisciplinary learning science* (1st ed., pp. 138–188). Oxford: Pergamon.

Guyatt, G. H. (1991). Evidence-based medicine. *ACP journal club*, *114*(2), A16-A16. https://doi.org/10.7326/ACPJC-1991-114-2-A16

*Hagell, A., & Spencer, L. (2004). An evaluation of an innovative audiotape method for keeping social care staff up to date with the latest research findings. *Child and Family Social Work*, *9*, 187–196. https://doi.org/10.1111/j.1365-2206.2004.00313.x

Harman, T., Bertrand, B., Greer, A., Pettus, A., Jennings, J., Wall-Bassett, E., & Babatunde, O. T. (2015). Case-based learning facilitates critical thinking in undergraduate nutrition education: Students describe the big picture. *Journal of the*

Academy of Nutrition and Dietetics, *115*(3), 378–388. https://doi.org/10.1016/j.jand.2014.09.003

Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement* (Reprinted.). London: Routledge.

Haynes, R. B., Devereaux, P. J., & Guyatt, G. H. (2002). Clinical expertise in the era of evidence-based medicine and patient choice. *ACP Journal Club*, *136*(2), A11-4.

Heiwe, S., Nilsson-Kajermo, K., Olsson, M., Gåfvels, C., Larsson, K., & Wengström, Y. (2013). Evidence-based practice among Swedish medical social workers. *Social Work in Health Care*, *52*(10), 947–958. https://doi.org/10.1080/00981389.2013.834029

Herbert, J. D. (2003). The science and practice of empirically supported treatments. *Behavior Modification*, *27*(3), 412–430. https://doi.org/10.1177/0145445503027003008

Hetmanek, A., Engelmann, K., Opitz, A., & Fischer, F. (2018). Beyond intelligence and domain knowledge: Scientific reasoning and argumentation as a set of cross-domain skills. In F. Fischer, C. A. Chinn, K. F. Engelmann, & J. Osborne (Eds.), *Scientific reasoning and argumentation: The roles of domain-specific and domain-general knowledge* (pp. 203–226). New York, London: Routledge.

Higgins, J. P. T., & Green, S. (Eds.). (2011). *Cochrane handbook of systematic reviews of interventions* (Version 5.1.0). Retrieved from http://handbook.cochrane.org

Hilbert, T. S., Renkl, A., Kessler, S., & Reiss, K. (2008). Learning to prove in geometry: Learning from heuristic examples and how it can be supported. *Learning and Instruction*, *18*(1), 54–65. https://doi.org/10.1016/j.learninstruc.2006.10.008

Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review*, *16*(3), 235–266. https://doi.org/10.1023/B:EDPR.0000034022.16470.f3

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, *42*(2), 99–107. https://doi.org/10.1080/00461520701263368

Hoadley, C., & van Heneghan, J. P. (2012). The learning sciences: Where they came from and what it means for instructional designers. In R. A. Reiser (Ed.), *Trends and issues in instructional design and technology* (3rd ed., pp. 53–63). Boston, Mass.: Pearson.

Hoffmann, J., & Engelkamp, J. (2017). *Lern- und Gedächtnispsychologie* (2., überarbeitete Auflage). *Springer-Lehrbuch*. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-662-49068-6

*Hohman, M., Pierce, P., & Barnett, E. (2015). Motivational interviewing: An evidence-based practice for improving student practice skills. *Journal of Social Work Education*, *51*(2), 287–297. https://doi.org/10.1080/10437797.2015.1012925

*Holmes, C. M. (2008). How can curriculum action research be used to promote post qualifying practice teaching award candidates' interest in research articles? *Social Work Education*, *27*(7), 695–709. https://doi.org/10.1080/02615470701538215

Holosko, M., & Leslie, D. R. (2001). Is social work a profession? The Canadian response. *Research on Social Work Practice*, *11*(2), 201–209. https://doi.org/10.1177/104973150101100206

Hong, S., & Yu, P. (2017). Comparison of the effectiveness of two styles of case-based learning implemented in lectures for developing nursing students' critical thinking ability: A randomized controlled trial. *International Journal of Nursing Studies*, *68*, 16–24. https://doi.org/10.1016/j.ijnurstu.2016.12.008

Horntvedt, M. E. T., Nordsteien, A., Fermann, T., & Severinsson, E. (2018). Strategies for teaching evidence-based practice in nursing education: A thematic literature review. *BMC Medical Education*, *18*(1), 172. https://doi.org/10.1186/s12909-018-1278-z

Horwitz, S. M., Hurlburt, M. S., Goldhaber-Fieber, J. D., Palinkas, L. A., Rolls-Reutz, J., Zhang, J., . . . Landsverk, J. (2014). Exploration and adoption of evidence-based practice by US child welfare agencies. *Children and Youth Services Review*, *39*, 147–152. https://doi.org/10.1016/j.childyouth.2013.10.004

Hüttemann, M. (2006). Evidence-based practice - ein Beitrag zur Professionalisierung Sozialer Arbeit? *Neue Praxis*, *36*(1), 36–51.

Ioannidis, J. P. A. (2016). The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. *The Milbank Quarterly*, *94*, 485–514. https://doi.org/10.1111/1468-0009.12210

James, S., Lampe, L., Behnken, S., & Schulz, D. (2019). Evidence-based practice and knowledge utilisation – a study of attitudes and practices among social workers in Germany. *European Journal of Social Work*, *22*(5), 763–777. https://doi.org/10.1080/13691457.2018.1469475

Jayaratne, S., & Levy, R. L. (1979). *Empirical clinical practice*. New York: Columbia Univ. Pr.

Johnson, B., & Christensen, L. (2013). *Educational research: Quantitative, qualitative, and mixed approaches* (5th ed.). Thousand Oaks, CA: SAGE Pub.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19*(4), 509–539. https://doi.org/10.1007/s10648-007-9054-3

Kalyuga, S., Rikers, R., & Paas, F. (2012). Educational implications of expertise reversal effects in learning and performance of complex cognitive and sensorimotor skills. *Educational Psychology Review*, *24*(2), 313–337. https://doi.org/10.1007/s10648-012-9195-x

Khacharem, A., Zoudji, B., & Kalyuga, S. (2015). Expertise reversal for different forms of instructional designs in dynamic visual representations. *British Journal of Educational Technology*, *46*(4), 756–767. https://doi.org/10.1111/bjet.12167

Khacharem, A., Zoudji, B., & Ripoll, H. (2013). Effect of presentation format and expertise on attacking-drill memorization in soccer. *Journal of Applied Sport Psychology*, *25*(2), 234–248. https://doi.org/10.1080/10413200.2012.718312

King, A. (1998). Transactive peer tutoring: Distributing cognition and metacognition. *Educational Psychology Review*, *10*(1), 57–74. https://doi.org/10.1023/A:1022858115001

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

Knight, C. (2013). Social workers' attitudes toward peer-reviewed literature: The evidence base. *Journal of Teaching in Social Work*, *33*(2), 177–195. https://doi.org/10.1080/08841233.2013.773955

Knight, C. (2015). Social work students' use of the peer-reviewed literature and engagement in evidence-based practice. *Journal of Social Work Education*, *51*(2), 250–269. https://doi.org/10.1080/10437797.2015.1012924

*Kobak, K. A., Mundt, J. C., & Kennard, B. (2015). Integrating technology into cognitive behavior therapy for adolescent depression: A pilot study. *Annals of General Psychiatry*, *14*, 37. https://doi.org/10.1186/s12991-015-0077-8

Kollar, I., & Fischer, F. (2019). Lehren und Unterrichten. In D. Urhahne, M. Dresel, & F. Fischer (Eds.), *Psychologie für den Lehrberuf* (pp. 333–352). Berlin: Springer Berlin.

Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration Scripts – A Conceptual Analysis. *Educational Psychology Review*, *18*(2), 159–185. https://doi.org/10.1007/s10648-006-9007-2

Kollar, I., Fischer, F., & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning and Instruction*, *17*(6), 708–721.

Kollar, I., Ufer, S., Reichersdorfer, E., Vogel, F., Fischer, F., & Reiss, K. (2014). Effects of collaboration scripts and heuristic worked examples on the acquisition of mathematical argumentation skills of teacher students with different levels of prior achievement. *Learning and Instruction*, *32*, 22–36. https://doi.org/10.1016/j.learninstruc.2014.01.003

Kolodner, J. L. (1992). An introduction to case-based reasoning. *Artificial Intelligence Review*, *6*(1), 3–34. https://doi.org/10.1007/BF00155578

Kolodner, J. L. (Ed.). (1993). *Case-based learning*. Boston, MA, s.l.: Springer US. https://doi.org/10.1007/978-1-4615-3228-6

Kolodner, J. L., Camp, P. J., Crismond, D., Fasse, B., Gray, J., Holbrook, J., . . . Ryan, M. (2003). Problem-based learning meets case-based reasoning in the middle-school science classroom: Putting Learning by Design(tm) into practice. *Journal of the Learning Sciences*, *12*(4), 495–547. https://doi.org/10.1207/S15327809JLS1204_2

Kolodner, J. L., Cox, M. T., & Gonzalez-Calero, P. A. (2005). Case-based reasoning-inspired approaches to education. *The Knowledge Engineering Review*, *20*, 299–303. https://doi.org/10.1017/S0269888906000634

Kolodner, J. L., Gray, J., & Fasse, B. B. (2003). Promoting transfer through case-based reasoning: rituals and practices in learning by design classro. *Cognitive Science Quaterly*, *3*, 183–232.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*, 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kopp, V., Stark, R., & Fischer, M. R. (2008). Fostering diagnostic knowledge through computer-supported, case-based worked examples: Effects of erroneous examples and feedback. *Medical Education*, *42*(8), 823–829. https://doi.org/10.1111/j.1365-2923.2008.03122.x

Kopp, V., Stark, R., Kühne-Eversmann, L., & Fischer, M. R. (2009). Do worked examples foster medical students' diagnostic knowledge of hyperthyroidism? *Medical Education*, *43*(12), 1210–1217. https://doi.org/10.1111/j.1365-2923.2009.03531.x

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121–1134.

Kuhn, D. (2007). Is direct instruction an answer to the right question? *Educational Psychologist*, *42*(2), 109–113. https://doi.org/10.1080/00461520701263376

Kuhn, D., Black, J., Keselman, A., & Kaplan, D. (2000). The development of cognitive skills to support inquiry learning. *Cognition and Instruction*, *18*, 495–523. https://doi.org/10.1207/S1532690XCI1804_3

Kyriakoulis, K., Patelarou, A., Laliotis, A., Wan, A. C., Matalliotakis, M., Tsiou, C., & Patelarou, E. (2016). Educational strategies for teaching evidence-based practice to undergraduate health students: Systematic review. *Journal of Educational Evaluation for Health Professions*, *13*, 34. https://doi.org/10.3352/jeehp.2016.13.34

Kyun, S., Kalyuga, S., & Sweller, J. (2013). The effect of worked examples when learning to write essays in english literature. *The Journal of Experimental Education*, *81*(3), 385–408. https://doi.org/10.1080/00220973.2012.727884

Lambers, H. (2018). *Theorien der Sozialen Arbeit: Ein Kompendium und Vergleich* (4., überarbeitete und erweiterte Auflage). *utb Soziale Arbeit: Vol. 4985*. Opladen, Toronto: Verlag Barbara Budrich.

Larsen, C. M., Terkelsen, A. S., Carlsen, A. M. F., & Kristensen, H. K. (2019). Methods for teaching evidence-based practice: A scoping review. *BMC Medical Education*, *19*(1), 259. https://doi.org/10.1186/s12909-019-1681-0

Lave, J. (1988). *Cognition in practice: Mind, mathematics and culture in everyday life*. New York, NY: Cambridge Univ. Press.

Lave, J., & Wenger, E. (1991). *Situated learning*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511815355

Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning. *Review of Educational Research*, *86*(3), 681–718. https://doi.org/10.3102/0034654315627366

*Leathers, S. J., Spielfogel, J. E., Blakey, J., Christian, E., & Atkins, M. S. (2016). The effect of a change agent on use of evidence-based mental health practices. *Administration and Policy in Mental Health*, *43*, 768–782. https://doi.org/10.1007/s10488-015-0694-1

*Leathers, S. J., & Strand, T. C. (2013). Increasing access to evidence-based practices and knowledge and attitudes. Research on Social Work Practice, 23, 669–679. https://doi:10.1177/1049731513491635

Lee, H., Plass, J. L., & Homer, B. D. (2006). Optimizing cognitive load for learning from computer-based science simulations. *Journal of Educational Psychology*, *98*(4), 902–913.

Lee, Y. H. (2015). Facilitating critical thinking using the C-QRAC collaboration script: Enhancing science reading literacy in a computer-supported collaborative learning environment. *Computers & Education*, *88*, 182–191. https://doi.org/10.1016/j.compedu.2015.05.004

Lee, Y. H. (2018). Scripting to enhance university students' critical thinking in flipped learning: Implications of the delayed effect on science reading literacy. *Interactive Learning Environments*, *26*(5), 569–582. https://doi.org/10.1080/10494820.2017.1372483

Liberati, A., Altman, D., Tetzlaff, J., Mulrow, C., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that

evaluate healthcare interventions: Explanation and elaboration. *PLoS medicine*, *6*(7). https://doi.org/10.1136/bmj.b2700

Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *2*(1), 53–70. https://doi.org/10.1111/j.1745-6916.2007.00029.x

Loibl, K., Roll, I., & Rummel, N. (2017). Towards a theory of when and how problem-solving followed by instruction supports learning. *Educational Psychology Review*, *29*(4), 693–715. https://doi.org/10.1007/s10648-016-9379-x

*Lopez, M. A., Osterberg, L. D., Jensen-Doss, A., & Rae, W. A. (2011). Effects of workshop training for providers under mandated use of an evidence-based practice. *Administration and Policy in Mental Health*, *38*, 301–312. https://doi.org/10.1007/s10488-010-0326-8

Lorains, M., Ball, K., & MacMahon, C. (2013). Expertise differences in a video decision-making task: Speed influences on performance. *Psychology of Sport and Exercise*, *14*(2), 293–297. https://doi.org/10.1016/j.psychsport.2012.11.004

Loyens, S. M. M., & Rikers, R. M. J. P. (2011). Instruction based on inquiry. In R. E. Mayer & P. A. Alexander (Eds.), *Educational psychology handbook series. Handbook of research on learning and instruction* (pp. 361–382). New York, NY: Routledge.

Luebbe, A. M., Radcliffe, A. M., Callands, T. A., Green, D., & Thorn, B. E. (2007). Evidence-based practice in psychology: Perceptions of graduate students in scientist-practitioner programs. *Journal of Clinical Psychology*, *63*(7), 643–655. https://doi.org/10.1002/jclp.20379

Macdonald, G. (1998). Promoting evidence-based practice in child protection. *Clinical Child Psychology and Psychiatry*, *3*(1), 71–85. https://doi.org/10.1177/1359104598031008

Mäkitalo, K., Weinberger, A., Häkkinen, P., Järvelä, S., & Fischer, F. (2005). Epistemic cooperation scripts in online learning environments: Fostering learning by reducing uncertainty in discourse? *Computers in Human Behavior*, *21*(4), 603–622. https://doi.org/10.1016/j.chb.2004.10.033

Mandl, H. & Friedrich, H. F. (2006). Handbuch Lernstrategien. Göttingen: Hogrefe.

*Martin, F., Waites, C., Hopp, F. P., Sobeck, J., & Agius, E. (2013). Enhancing aging services through evidence-based health promotion: A training for service providers. *Journal of Evidence-Informed Social Work*, *10*, 482–493. https://doi.org/10.1080/15433714.2012.760296

*Matthieu, M. M., Carter, L., Casner, R. W., & Edmond, T. E. (2016). Training outcomes of field instructors in the evidence-based practice process model. *Journal of Teaching in Social Work*, *36*, 477–489. https://doi.org/10.1080/08841233.2016.1242524

Mayer, R. E. (2002). Rote versus meaningful learning. *Theory Into Practice*, *41*(4), 226–232. https://doi.org/10.1207/s15430421tip4104_4

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist*, *59*(1), 14–19. https://doi.org/10.1037/0003-066X.59.1.14

Mayer, R. E., Steinhoff, K., Bower, G., & Mars, R. (1995). A generative theory of textbook design: Using annotated illustrations to foster meaningful learning of science text.

Educational Technology Research and Development, *43*(1), 31–41. https://doi.org/10.1007/BF02300480

McGaugh, J. L. (2013). Making lasting memories: Remembering the significant. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(sup2), 10402–10407. https://doi.org/10.1073/pnas.1301209110

McGilly, K. (1998). *Classroom lessons: Integrating cognitive theory and classroom practice* (4. print). *A Bradford book.*

McNeece, C. A., & Thyer, B. A. (2004). Evidence-based practice and social work. *Journal of Evidence-Based Social Work*, *1*(1), 7–25. https://doi.org/10.1300/J394v01n01_02

Micheel, H.-G. (2010). Die Leistungsfähigkeit empirischer Sozialforschung im Kontext sozialpädagogischer Praxis. In H. U. Otto, A. Polutta, & H. Ziegler (Eds.), *What works - welches Wissen braucht die soziale Arbeit?: Zum Konzept evidenzbasierter Praxis* (pp. 151–163). Opladen, Farmington Hills, MI: Verlag Barbara Budrich.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Morago, P. (2010). Dissemination and implementation of evidence-based practice in the social services: A UK survey. *Journal of Evidence-Informed Social Work*, *7*(5), 452–465. https://doi.org/10.1080/15433714.2010.494973

Mullen, E. J. (2016). Reconsidering the 'idea' of evidence in evidence-based policy and practice. *European Journal of Social Work*, *19*(3-4), 310–335. https://doi.org/10.1080/13691457.2015.1022716

Mullen, E. J., & Streiner, D. L. (2004). The evidence for and against evidence-based practice. *Brief Treatment and Crisis Intervention*, *4*(2), 111–121. https://doi.org/10.1093/brief-treatment/mhh009

Munro, E. (2002). The role of theory in social work research: A further contribution to the debate. *Journal of Social Work Education, 38*(3), 461–470

Murphy, A., & McDonald, J. (2004). Power, status and marginalisation: Rural social workers and evidence-based practice in multidisciplinary teams. *Australian Social Work*, *57*(2), 127–136. https://doi.org/10.1111/j.1447-0748.2004.00127.x

National Association of Social Workers (2017). NASW code of ethics. Retrieved from https://www.socialworkers.org/about/ethics/code-of-ethics/code-of-ethics-english

Neuman, Y. (2003). Go ahead, prove that god does not exist! On high school students' ability to deal with fallacious arguments. *Learning and Instruction*, *13*(4), 367–380. https://doi.org/10.1016/S0959-4752(02)00011-7

Neuman, Y., Glassner, A., & Weinstock, M. (2004). The effect of a reason's truth-value on the judgment of a fallacious argument. *Acta Psychologica*, *116*(2), 173–184. https://doi.org/10.1016/j.actpsy.2004.01.003

Neuman, Y., Weinstock, M. P., & Glasner, A. (2006). The effect of contextual factors on the judgement of informal reasoning fallacies. *Quarterly Journal of Experimental Psychology (2006)*, *59*(2), 411–425. https://doi.org/10.1080/17470210500151436

Neuman, Y., & Weizman, E. (2003). The role of text representation in students' ability to identify fallacious arguments. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *56*(5), 849–864. https://doi.org/10.1080/02724980244000666

Nievelstein, F., van Gog, T., van Dijck, G., & Boshuizen, H. P. A. (2013). The worked example and expertise reversal effect in less structured tasks: Learning to reason about legal cases. *Contemporary Educational Psychology*, *38*(2), 118–125. https://doi.org/10.1016/j.cedpsych.2012.12.004

Noroozi, O., Kirschner, P. A., Biemans, H. J.A., & Mulder, M. (2018). Promoting argumentation competence: Extending from first- to second-order scaffolding through adaptive fading. *Educational Psychology Review*, *30*(1), 153–176. https://doi.org/10.1007/s10648-017-9400-z

Noroozi, O., Weinberger, A., Biemans, H. J.A., Mulder, M., & Chizari, M. (2012). Argumentation-based computer supported collaborative learning (ABCSCL): A synthesis of 15 years of research. *Educational Research Review*, *7*(2), 79–106. https://doi.org/10.1016/j.edurev.2011.11.006

Nugent, W. R. (2010). Probability and sampling. In B. A. Thyer (Ed.), *The handbook of social work research methods* (2nd ed., pp. 37–50). Los Angeles: SAGE.

Okpych, N. J., & Yu, J. L. H. (2014). A historical analysis of evidence-based practice in social work: The unfinished journey toward an empirically grounded profession. *Social Service Review*, *88*(1), 3–58. https://doi.org/10.1086/674969

Oksa, A., Kalyuga, S., & Chandler, P. (2010). Expertise reversal effect in using explanatory notes for readers of Shakespearean text. *Instructional Science*, *38*(3), 217–236. https://doi.org/10.1007/s11251-009-9109-6

Otto, H. U., & Ziegler, H. (2008). The notion of causal impact in evidence-based social work: An introduction to the special issue on what works? *Research on Social Work Practice*, *18*(4), 273–277. https://doi.org/10.1177/1049731507313997

Otto, H. U., Polutta, A., & Ziegler, H. (Eds.). (2010). *What works - welches Wissen braucht die soziale Arbeit?: Zum Konzept evidenzbasierter Praxis*. Opladen, Farmington Hills, MI: Verlag Barbara Budrich.

Owens, P., & Sweller, J. (2008). Cognitive load theory and music instruction. *Educational Psychology*, *28*(1), 29–45. https://doi.org/10.1080/01443410701369146

Paas, F., & van Gog, T. (2006). Optimising worked example instruction: Different ways to increase germane cognitive load. *Learning and Instruction*, *16*(2), 87–91. https://doi.org/10.1016/j.learninstruc.2006.02.004

Paas, F., & van Merrienboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology*, *86*(1), 122–133. https://doi.org/10.1037/0022-0663.86.1.122

Pace, R., Pluye, P., Bartlett, G., Macaulay, A. C., Salsberg, J., Jagosh, J., & Seller, R. (2012). Testing the reliability and efficiency of the pilot Mixed Methods Appraisal Tool (MMAT) for systematic mixed studies review. *International Journal of Nursing Studies*, *49*, 47–53. https://doi.org/10.1016/j.ijnurstu.2011.07.002

Palinscar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, *1*(2), 117–175. https://doi.org/10.1207/s1532690xci0102_1

*Parrish, D. E., & Rubin, A. (2011). An effective model for continuing education training in evidence-based practice. *Research on Social Work Practice*, *21*, 77–87. https://doi.org/10.1177/1049731509359187

Parrish, D. E., & Rubin, A. (2012). Social workers' orientations toward the evidence-based practice process: A comparison with psychologists and licensed marriage and family therapists. *Social Work*, *57*(3), 201–210. https://doi.org/10.1093/sw/sws016

Patelarou, A., Kyriakoulis, K. G., Stamou, A. A., Laliotis, A., Sifaki-Pistolla, D., Matalliotakis, M., . . . Patelarou, E. (2017). Approaches to teach evidence-based practice among health professionals: An overview of the existing evidence. *Advances in Medical Education and Practice*, *8*, 455–464. https://doi.org/10.2147/AMEP.S134475

*Peterson, S. M., Phillips, A., Bacon, S. I., & Machunda, Z. (2011). Teaching evidence-based practice at the BSW level: An effective capstone project. *Journal of Social Work Education*, *47*, 509–524. https://doi.org/10.5175/JSWE.2011.200900129

Pfister, H. R., & Mühlpfordt, M. (2002). Supporting discourse in a synchronous learning environment: The learning protocol approach. In G. Stahl (Ed.), *Proceedings of the conference on computer supported collaborative learning (CSCL)* (pp. 581–591). Hillsdale: Erlbaum.

Pignotti, M., & Thyer, B. A. (2012). Novel unsupported and empirically supported therapies: Patterns of usage among licensed clinical social workers. *Behavioural and Cognitive Psychotherapy*, *40*(3), 331–349. https://doi.org/10.1017/S135246581100052X

Pluye, P., Gagnon, M. P., Griffiths, F., & Johnson-Lafleur, J. (2009). A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in mixed studies reviews. *International Journal of Nursing Studies*, *46*, 529–546. https://doi.org/10.1016/j.ijnurstu.2009.01.009

Pluye, P., Robert, E., Cargo, M., Bartlett, G., O'Cathain, A., Griffiths, F., Boardman, F., Gagnon, M. P., & Rousseau, M. C. (2011). Proposal: A mixed methods appraisal tool for systematic mixed studies reviews. Retrieved from http://mixedmethodsappraisaltoolpublic.pbworks.com

Pociask, F. D., & Morrison, G. R. (2008). Controlling split attention and redundancy in physical therapy instruction. *Educational Technology Research and Development*, *56*(4), 379–399. https://doi.org/10.1007/s11423-007-9062-5

Pope, N. D., Rollins, L., Chaumba, J., & Risler, E. (2011). Evidence-based practice knowledge and utilization among social workers. *Journal of Evidence-Informed Social Work*, *8*(4), 349–368. https://doi.org/10.1080/15433710903269149

Popil, I. (2011). Promotion of critical thinking by using case studies as teaching method. *Nurse education today*, *31*(2), 204–207. https://doi.org/10.1016/j.nedt.2010.06.002

Postma, T. C., & White, J. G. (2016). Developing integrated clinical reasoning competencies in dental students using scaffolded case-based learning - empirical evidence. *European Journal of Dental Education: Official Journal of the Association for Dental Education in Europe*, *20*(3), 180–188. https://doi.org/10.1111/eje.12159

Profetto-McGrath, J. (2005). Critical thinking and evidence-based practice. *Journal of Professional Nursing: Official Journal of the American Association of Colleges of Nursing*, *21*(6), 364–371. https://doi.org/10.1016/j.profnurs.2005.10.002

Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., . . . Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *Journal of the Learning Sciences*, *13*(3), 337–386. https://doi.org/10.1207/s15327809jls1303_4

Ramasamy, S., & Fung, H. P. (2015, October). *The analysis of informal reasoning fallacy and critical thinking disposition among malaysian undergraduates*. Proceedings of the Aeu International Research Conference, Kuala Lumpur.

Reisslein, J., Atkinson, R. K., Seeling, P., & Reisslein, M. (2006). Encountering the expertise reversal effect with a computer-based environment on electrical circuit analysis. *Learning and Instruction*, *16*(2), 92–103. https://doi.org/10.1016/j.learninstruc.2006.02.008

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive science*, *21*(1), 1–29. https://doi.org/10.1207/s15516709cog2101_1

Renkl, A. (2005). The worked-out examples principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 229–246). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511816819.016

Renkl, A. (2009). Lehren und Lernen. In R. Tippelt & B. Schmidt (Eds.), *Handbuch Bildungsforschung* (2nd ed., pp. 737–751). Wiesbaden: VS Verlag für Sozialwissenschaften / GWV Fachverlage GmbH Wiesbaden.

Renkl, A. (2011). Instruction based on examples. In R. E. Mayer & P. A. Alexander (Eds.), *Educational psychology handbook series. Handbook of research on learning and instruction* (pp. 272–295). New York, NY: Routledge.

Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, *38*(1), 1–37. https://doi.org/10.1111/cogs.12086

Renkl, A. (2017). Instruction based on examples. In R. E. Mayer & P. A. Alexander (Eds.), *Educational psychology handbook series. Handbook of research on learning and instruction* (pp. 325–348). New York, NY, London: Routledge.

Renkl, A., Hilbert, T., & Schworm, S. (2009). Example-based learning in heuristic domains: A cognitive load theory account. *Educational Psychology Review*, *21*(1), 67–78. https://doi.org/10.1007/s10648-008-9093-4

Renkl, A., Mandl, H., & Gruber, H. (1996). Inert knowledge: Analyses and remedies. *Educational Psychologist*, *31*(2), 115–121. https://doi.org/10.1207/s15326985ep3102_3

*Ronen, T. (2005). Students' evidence-based practice intervention for children with oppositional defiant disorder. *Research on Social Work Practice*, *15*(3), 165–179. https://doi.org/10.1177/1049731504271604

Rosen, A., Proctor, E. K., & Staudt, M. M. (1999). Social work research and the quest for effective practice. *Social Work Research*, *23*(1), 4–14. https://doi.org/10.1093/swr/23.1.4

Rourke, A., & Sweller, J. (2009). The worked-example effect using ill-defined problems: Learning to recognise designers' styles. *Learning and Instruction*, *19*(2), 185–199. https://doi.org/10.1016/j.learninstruc.2008.03.006

Rousmaniere, T., Goodyear, R. K., Miller, S. D., & Wampold, B. E. (Eds.). (2017). *The cycle of excellence: Using deliberate practice to improve supervision and training*. Chichester, UK, Hoboken, NJ: John Wiley & Sons.

Rubin, A., & Parrish, D. (2007). Views of evidence-based practice among faculty in master of social work programs: A national survey. *Research on Social Work Practice*, *17*(1), 110–122. https://doi.org/10.1177/1049731506293059

Rubin, A., & Parrish, D. (2010). Development and validation of the Evidence-Based Practice Process Assessment Scale: Preliminary findings. *Research on Social Work Practice*, *20*(6), 629–640. https://doi.org/10.1177/1049731508329420

Rummel, N., Spada, H., & Hauser, S. (2009). Learning to collaborate while being scripted or by observing a model. *International Journal of Computer-Supported Collaborative Learning*, *4*(1), 69–92. https://doi.org/10.1007/s11412-008-9054-4

Ruzafa-Martínez, M., López-Iborra, L., Armero Barranco, D., & Ramos-Morcillo, A. J. (2016). Effectiveness of an evidence-based practice (EBP) course on the EBP competence of undergraduate nursing students: A quasi-experimental study. *Nurse Education Today*, *38*, 82–87. https://doi.org/10.1016/j.nedt.2015.12.012

*Sacco, P., Ting, L., Crouch, T. B., Emery, L., Moreland, M., Bright, C., . . . DiClemente, C. (2017). SBIRT training in social work education: Evaluating change using standardized patient simulation. *Journal of Social Work Practice in the Addictions*, *17*, 150–168. https://doi.org/10.1080/1533256X.2017.1302886

Sackett, D. L., Richardson, W. S., Rosenberg, W., & Haynes, R. B. (1997). *Evidence-based medicine: How to practice and teach EBM*. New York, NY: Churchill Livingstone.

Sackett, D. L., & Rosenberg, W. (1995). On the need for evidence-based medicine. *Health Economics*, *4*(4), 249–254. https://doi.org/10.1002/hec.4730040401

Sackett, D. L., Rosenberg, W., Gray, J. A., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *BMJ : British Medical Journal*, *312*(7023), 71–72.

Sackett, D. L., Straus, S. E., Richardson, W. S., Rosenberg, W., & Haynes R. B. (2000). *Evidence-based medicine: How to practice and teach EBM* (2. ed.). Edinburgh: Churchill Livingstone.

*Salcido, R. M. (2008). Incorporating evidence-based practice into the macro practice curriculum. *Journal of Evidence-Informed Social Work*, *5*(3-4), 623–645. https://doi.org/10.1080/15433710802084359

Salden, R. J. C. M., Aleven, V., Schwonke, R., & Renkl, A. (2010). The expertise reversal effect and worked examples in tutored problem-solving. *Instructional Science*, *38*(3), 289–307. https://doi.org/10.1007/s11251-009-9107-8

*Salloum, A., & Smyth, K. (2013). Clinicians' experiences of a podcast series on implementing a manualized treatment. *Journal of Technology in Human Services*, *31*(1), 71–83. https://doi.org/10.1080/15228835.2012.738382

Satterfield, J. M., Spring, B., Brownson, R. C., Mullen, E. J., Newhouse, R. P., Walker, B. B., & Whitlock, E. P. (2009). Toward a transdisciplinary model of evidence-based practice. *The Milbank Quarterly*, *87*(2), 368–390. https://doi.org/10.1111/j.1468-0009.2009.00561.x

Sawyer, R. K. (Ed.). (2009). *The Cambridge handbook of the learning sciences* (Reprinted.). Cambridge: Cambridge Univ. Press.

Schank, R. C. (1999). *Dynamic memory revisited*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511527920

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. *Artificial Intelligence Series*. Hoboken: Taylor and Francis.

Schellens, T., van Keer, H., Wever, B. de, & Valcke, M. (2009). Tagging thinking types in asynchronous discussion groups: Effects on critical thinking. *Interactive Learning Environments*, *17*(1), 77–94. https://doi.org/10.1080/10494820701651757

Schmidt, H. G., Loyens, S. M. M., van Gog, T., & Paas, F. (2007). Problem-based learning is compatible with human cognitive architecture: Commentary on Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, *42*(2), 91–97. https://doi.org/10.1080/00461520701263350

Schmidt, H. G., van der Molen, H. T., Te Winkel, W. W. R., & Wijnen, W. H. F. W. (2009). Constructivist, problem-based learning does work: A meta-analysis of curricular comparisons involving a single medical school. *Educational Psychologist*, *44*(4), 227–249. https://doi.org/10.1080/00461520903213592

Schwonke, R., Renkl, A., Krieg, C., Wittwer, J., Aleven, V., & Salden, R. J. C. M. (2009). The worked-example effect: Not an artefact of lousy control conditions. *Computers in Human Behavior*, *25*(2), 258–266. https://doi.org/10.1016/j.chb.2008.12.011

Schworm, S., & Renkl, A. (2007). Learning argumentation skills through the use of prompts for self-explaining examples. *Journal of Educational Psychology*, *99*(2), 285–296. https://doi.org/10.1037/0022-0663.99.2.285

Scurlock-Evans, L., & Upton, D. (2015). The role and nature of evidence: A systematic review of social workers' evidence-based practice orientation, attitudes, and implementation. *Journal of Evidence-Informed Social Work*, *12*(4), 369–399. https://doi.org/10.1080/15433714.2013.853014

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

Silverman, W. A. (1980). *Retrolental fibroplasia: A modern parable*. New York: Grune & Stratton.

Slavin, R. E. (2008). Evidence-based reform in education: What will it take? *European Educational Research Journal*, *7*(1), 124–128. https://doi.org/10.2304/eerj.2008.7.1.124

Slavin, R. E. (2018). *Educational psychology: Theory and practice* (Twelfth edition). NY, NY: Pearson.

*Smith, C. A., Cohen-Callow, A., Harnek-Hall, D. M., & Hayward, R. A. (2007). Impact of a foundation-level MSW research course on students' critical appraisal skills. *Journal of Social Work Education*, *43*(3), 481–495.

Snibsøer, A. K., Ciliska, D., Yost, J., Graverholt, B., Nortvedt, M. W., Riise, T., & Espehaug, B. (2018). Self-reported and objectively assessed knowledge of evidence-based practice terminology among healthcare students: A cross-sectional study. *PloS One*, *13*(7), e0200313. https://doi.org/10.1371/journal.pone.0200313

Sommerfeld, P., & Hüttemann, M. (Eds.). (2007). *Grundlagen der sozialen Arbeit: Vol. 17. Evidenzbasierte Soziale Arbeit: Nutzung von Forschung in der Praxis*. Baltmannsweiler: Schneider Verl. Hohengehren.

Souto, R. Q., Khanassov, V., Hong, Q. N., Bush, P. L., Vedel, I., & Pluye, P. (2015). Systematic mixed studies reviews: Updating results on the reliability and efficiency of the Mixed

Methods Appraisal Tool. *International Journal of Nursing Studies*, *52*, 500–501. https://doi.org/10.1016/j.ijnurstu.2014.08.010

Spensberger, F., Kollar, I., & Pankofer, S. (2019). Effects of worked examples and external scripts on social work students' fallacy recognition skills. Open Science Framework. Retrieved from osf.io/5p86x

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74*(11), 1–29. https://doi.org/10.1037/h0093759

Stark, R., Kopp, V., & Fischer, M. R. (2011). Case-based learning with worked examples in complex domains: Two experimental studies in undergraduate medical education. *Learning and Instruction*, *21*(1), 22–33. https://doi.org/10.1016/j.learninstruc.2009.10.001

Staub-Bernasconi, S. (2018). *Soziale Arbeit als Handlungswissenschaft: Soziale Arbeit auf dem Weg zu kritischer Professionalität* (2., vollständig überarbeitete u. aktualisierte Ausgabe). *utb Soziale Arbeit: Vol. 2786*. Opladen, Toronto, Leverkusen: Verlag Barbara Budrich; UTB.

Stegmann, K., Weinberger, A., & Fischer, F. (2007). Facilitating argumentative knowledge construction with computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning*, *2*(4), 421–447. https://doi.org/10.1007/s11412-007-9028-y

Straus, S. E., & McAlister, F. A. (2000). Evidence-based medicine: A commentary on common criticisms. *Canadian Medial Association Journal*, *163*(7), 837–841.

Straus, S. E., Richardson, W. S., Glasziou, P., & Haynes, R. B. (2005). *Evidence-based medicine: How to practice and teach EBM* (3. ed.). Edinburgh: Elsevier/Churchill Livingstone.

Straus, S. E., Richardson, W. S., Glasziou, P., & Haynes, R. B. (2011). *Evidence-based medicine: How to practice and teach it* (4. ed., reprinted.). Edinburgh: Churchill Livingstone Elsevier.

*Straussner, S. L. A., Naegle, M. A., Gillespie, C., Wolkstein, E., Donath, R., & Azmitia, E. C. (2006). The SATOL project: An interdisciplinary model of technology transfer for research-to-practice in clinical supervision for addiction treatment. *Journal of Evidence-Informed Social Work*, *3*, 39–54. https://doi.org/10.1300/J394v03n03_04

Sur, R. L., & Dahm, P. (2011). History of evidence-based medicine. *Indian Journal of Urology: IJU: Journal of the Urological Society of India*, *27*(4), 487–489. https://doi.org/10.4103/0970-1591.91438

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory* (1st ed.). *Explorations in the Learning Sciences, Instructional Systems and Performance Technologies: Vol. 1*. New York, NY: Springer Science+Business Media LLC. https://doi.org/10.1007/978-1-4419-8126-4

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, *12*(3), 185–233. https://doi.org/10.1207/s1532690xci1203_1

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem-solving in learning algebra. *Cognition and Instruction*, *2*(1), 59–89. https://doi.org/10.1207/s1532690xci0201_3

Sweller, J., Kirschner, P. A., & Clark, R. E. (2007). Why minimally guided teaching techniques do not work: A reply to commentaries. *Educational Psychologist*, *42*(2), 115–121. https://doi.org/10.1080/00461520701263426

Sweller, J., & Paas, F. (2017). Should self-regulated learning be integrated with cognitive load theory? A commentary. *Learning and Instruction*, *51*, 85–89. https://doi.org/10.1016/j.learninstruc.2017.05.005

Tabak, I. (2004). Synergy: A complement to emerging patterns of distributed scaffolding. *Journal of the Learning Sciences*, *13*(3), 305–335. https://doi.org/10.1207/s15327809jls1303_3

Tarmizi, R. A., & Sweller, J. (1988). Guidance during mathematical problem-solving. *Journal of Educational Psychology*, *80*(4), 424–436. https://doi.org/10.1037/0022-0663.80.4.424

Tarnoff, J. (2010). *An investigation of the role of confirmation bias in the evaluation of informal reasoning fallacies*. Temple University.

Task Force. (1993). Task force on the promotion and dissemination of psychological procedures. Retrieved from http://www.div12.org/sites/default/files/InitialReportOfTheChamblessTaskForce.pdf

Teater, B., & Chonody, J. M. (2018). Identifying as an evidence-based social worker: The influence of attitudes, access, confidence, and education. *Social Work Education*, *37*(4), 442–457. https://doi.org/10.1080/02615479.2017.1421161

*Tennille, J., Solomon, P., Brusilovskiy, E., & Mandell, D. (2016). Field instructors extending EBP learning in dyads (FIELD): Results of a pilot randomized controlled

trial. *Journal of the Society for Social Work and Research*, *7*(1), 1–22. https://doi.org/10.1086/684020

Thyer, B. A. (2001). Research on social work practice does not benefit from blurry theory: A response to Tomi Gomory. *Journal of Social Work Education*, *37*(1), 51–66.

Thyer, B. A. (2013). Evidence-based practice or evidence-guided practice: A rose by any other name would smell as sweet. *Families in Society: The Journal of Contemporary Social Services*, *94*(2), 79–84. https://doi.org/10.1606/1044-3894.4283

Thyer, B. A. (2015). Preparing current and future practitioners to integrate research in real practice settings. *Research on Social Work Practice*, *25*(4), 463–472. https://doi.org/10.1177/1049731514538105

Thyer, B. A., & Myers, L. L. (2011). The quest for evidence-based practice: A view from the United States. *Journal of Social Work*, *11*(1), 8–25. https://doi.org/10.1177/1468017310381812

Tian, J. H., Liu, A. P., Yang, K. H., & Shen, X. (2013). The effectiveness of problem-based learning on evidence-based medicine: A double-blind randomized trial. *Indian Journal of Pharmaceutical Education and Research*, *47*, 14–18.

Toulmin, S. (2003). *The uses of argument* (Second edition). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511840005

Van der Zwet, R. J. M., Kolmer, D. M. B. G., & Schalk, R. (2016). Social workers' orientation toward the evidence-based practice process: A dutch survey. *Research on Social Work Practice*, *26*(6), 712–722. https://doi.org/10.1177/1049731514540340

Van der Zwet, R. J. M., Kolmer, D. M. B. G., Schalk, R., & van Regenmortel, T. (2019). Views and attitudes towards evidence-based practice in a dutch social work organization.

Journal of evidence-based social work, *16*(3), 245–260. https://doi.org/10.1080/23761407.2019.1584071

Van der Zwet, R. J. M., Weling, J., Kolmer, D. M. B. G., & Schalk, R. (2017). Exploring MSW students' and social workers' orientation toward the evidence-based practice process. *Social Work Education*, *36*(1), 75–87. https://doi.org/10.1080/02615479.2016.1266321

Van Gog, T., Kester, L., & Paas, F. (2011). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology*, *36*, 212–218. https://doi.org/10.1016/j.cedpsych.2010.10.004

Van Merrienboer, J. J. G. (2013). Perspectives on problem-solving and instruction. *Computers & Education*, *64*, 153–160. https://doi.org/10.1016/j.compedu.2012.11.025

Van Merrienboer, J. J. G., Kirschner, P. A., & Kester, L. (2003). Taking the load off a learner's mind: Instructional design for complex learning. *Educational Psychologist*, *38*(1), 5–13. https://doi.org/10.1207/S15326985EP3801_2

Veenman, M. V. J., van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning*, *1*(1), 3–14. https://doi.org/10.1007/s11409-006-6893-0

Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-cognitive scaffolding with computer-supported collaboration scripts: A meta-analysis. *Educational Psychology Review*, *29*(3), 477–511. https://doi.org/10.1007/s10648-016-9361-7

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, Massachusetts, London: Harvard University Press.

Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes*: Harvard University Press.

Walton, D. N. (1995). *A pragmatic theory of fallacy*. *Studies in rhetoric and communication.*

Walton, D. N. (2009). Argumentation theory: A very short introduction. In G. Simari & I. Rahwan (Eds.), *Argumentation in artificial intelligence* (Vol. 13, pp. 1–22). Boston, MA: Springer-Verlag US. https://doi.org/10.1007/978-0-387-98197-0_1

Wang, X., Kollar, I., & Stegmann, K. (2017). Adaptable scripting to foster regulation processes and skills in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, *12*(2), 153–172. https://doi.org/10.1007/s11412-017-9254-x

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond "p < 0.05". *The American Statistician*, *73*, 1–19. https://doi.org/10.1080/00031305.2019.1583913

Webb, S. (2001). Some considerations on the validity of evidence-based practice in social work. *British Journal of Social Work*, *31*(1), 57–79. https://doi.org/10.1093/bjsw/31.1.57

*Webber, M., Currin, L., Groves, N., Hay, D., & Fernando, N. (2010). Social workers can e-learn: Evaluation of a pilot post-qualifying e-learning course in research methods and critical appraisal skills for social workers. *Social Work Education*, *29*(1), 48–66. https://doi.org/10.1080/02615470902838745

Wecker, C., Kollar, I., Fischer, F., & Prechtl, H. (2010). Fostering online search competence and domain-specific knowledge in inquiry classrooms: Effects of continuous and

fading collaboration scripts. In K. Gomez, L. Lyons, & J. Radinsky (Eds.), *Learning in the disciplines: Proceedings of the 9th international conference of the learning sciences—ICLS 2010* (pp. 810–817). Chicago.

Weinstock, M., Neuman, Y., & Tabak, I. (2004). Missing the point or missing the norms? Epistemological norms as predictors of students' ability to identify fallacious arguments. *Contemporary Educational Psychology*, *29*, 77–94. https://doi.org/10.1016/S0361-476X(03)00024-9

Weissman, M. M., Verdeli, H., Gameroff, M. J., Bledsoe, S. E., Betts, K., Mufson, L., . . . Wickramaratne, P. (2006). National survey of psychotherapy training in psychiatry, psychology, and social work. *Social Work*, *63*(8), 925–934. https://doi.org/10.1001/archpsyc.63.8.925

Wike, T. L., Bledsoe, S. E., Manuel, J. I., Despard, M., Johnson, L. V., Bellamy, J. L., & Killian-Farrell, C. (2014). Evidence-based practice in social work: Challenges and opportunities for clinicians and organizations. *Clinical Social Work Journal*, *42*(2), 161–170. https://doi.org/10.1007/s10615-014-0492-3

Wilson, B. G., & Myers, K. (2000). Situated cognition in theoretical and practical context. In D. H. Jonassen & Land S. M. (Eds.), *Theoretical foundations of learning environments* (pp. 57–88). Mahwah, NJ: Erlbaum.

Wittwer, J., & Renkl, A. (2010). How effective are instructional explanations in example-based learning? A meta-analytic review. *Educational Psychology Review*, *22*, 393–409. https://doi.org/10.1007/s10648-010-9136-5

*Wong, R. (2017). Evaluating a teaching module on ethically responsible evidence-based practice decision making in an advanced micro practice course. *Journal of Social Work Education*, *53*, 240–259. https://doi.org/10.1080/10437797.2016.1260505

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem-solving. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *17*(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Yaffe, J. (2013). Guest editorial—where's the evidence for social work education? *Journal of Social Work Education*, *49*(4), 525–527. https://doi.org/10.1080/10437797.2013.820582

Zoelch, C., Berner, V. D., & Thomas, J. (2019). Gedächtnis und Wissenserwerb. In D. Urhahne, M. Dresel, & F. Fischer (Eds.), *Psychologie für den Lehrberuf.* Berlin: Springer Berlin.

Academic Integrity Statement

Spensberger, Florian

_____

Name, Vorname
Last name, first name

Ich versichere, dass ich die an der Fakultät für Psychologie und Pädagogik der Ludwig-Maximilians-Universität München zur Dissertation eingereichte Arbeit mit dem Titel:
I assert that the thesis I submitted to the Faculty of Psychology and Pedagogy of the Ludwig-Maximilian-Universität-München under the title:

"Fostering Evidence-based Practice and Argument Evaluation in Social Work"

selbst verfasst, alle Teile eigenständig formuliert und keine fremden Textteile übernommen habe, die nicht als solche gekennzeichnet sind. Kein Abschnitt der Doktorarbeit wurde von einer anderen Person formuliert, und bei der Abfassung wurden keine anderen als die in der Abhandlung aufgeführten Hilfsmittel benutzt.
is written by myself, I have formulated all parts independently and I have not taken any texts components of others without indicating them. No formulation has been made by someone else and I have not used any sources other than indicated in the thesis.

Ich erkläre, dass ich an keiner anderen Stelle einen Antrag auf Zulassung zur Promotion gestellt oder bereits einen Doktortitel auf der Grundlage des vorgelegten Studienabschlusses erworben und mich auch nicht einer Doktorprüfung erfolglos unterzogen habe.
I assert I have not applied anywhere else for a doctoral degree nor have I obtained a doctor title on the basis of my present studies or failed a doctoral examination.

München, 04.05.2020

_____

Ort, Datum
Place, Date

Florian Spensberger

_____

Unterschrift Doktorand
Signature of the doctoral candidate