Appendix A

Table A1

Description of Eligibility Criteria

Eligibility Criteria	Operationalization
Empirical	A publication is considered ‘empirical’ if it reports on at least one study that draws on primary data, collected through methods like interviews, questionnaires, and the like.
Sample	The sample of the intervention must include at least one social worker (e.g. clinical supervisor, staff clinicians, field instructor, probation officer, addiction counsellor) and/ or social work student (Bachelor of Social Work (BSW), Master of Social Work (MSW), doctoral student (PhD)).
Intervention	An intervention study is considered to be a study which examines effects of an intervention on a sample. The study could be experimental, quasi-experimental or descriptive and could follow a quantitative, a qualitative or a mixed-methods approach. Therefore, we do not apply the term ‘effect’ solely to quantitative results. An effect could, for example, also be derived from an interviewees’ feedback that an intervention increased his knowledge. An intervention could be a training, workshop, course, class, task, method, teaching or instruction and the like. The primary focus of the intervention must be the facilitation of EBP or ESTs. A primary focus of the study must be the evaluation of the learners’ outcomes from the intervention.
EBP	An intervention study is considered to address the process of EBP and/or ESTs if the intervention teaches (1) the process of EBP or at least the search for best research evidence and/or the critical appraisal of research evidence and/or the usage of/ knowledge about research evidence and/or (2) knowledge about/the use of an EST. Following a rather inclusive approach, a treatment was considered as an EST, if it´s effectiveness was evaluated in at least one empirical study. An intervention study is not automatically considered to address the process of EBP or ESTs, just because the article claims that the intervention itself (workshop, course, etc.) is evidence-based. Furthermore, it is not automatically considered to address the process of EBP or ESTs, just by claiming that a certain program, theory, etc. which is under investigation is evidence-based, without providing further evidence for this claim.

Note. EBP = Process of Evidence-based Practice, EST/s = Empirically Supported Intervention/s

Table A2

Extracted Outcome Data

Author and Year	**Coded Effect: Measure Instrument** (Reliability/Validity)	**Design, Main Findings** (Coded Effect (Construct Measured in Study[a]): Evidence)
Ager et al. 2005	TDK: MCT (α =.87, r = .61) TPK: Vignettes (ICC = .68) PPK: LSQ (α = .65, r = .48 - .54) M: LSQ (α =.82, r = .69)	**3 groups pre vs. post vs. follow-up (range, x̄ (s)):** **(+) TDK:** 0-11, 4.7 (3.5) vs. 8.4 (2.1) vs. 8.5 (1.8) (EG1) 0-11, 4.8 (3.4) vs. 5.4 (3.5) vs. 8.2 (2.4) (EG2[b]) 0-11, 3.9 (3.2) vs. 4.4 (3.5) vs. 4.2 (3.4) CG) **(+) TPK:** 1-3, 1.7 (0.7) vs. 2.3 (0.7) vs. 2.3 (0.7) (EG1) 1-3, 1.7 (0.7) vs. 2.0 (0.7) vs. 2.4 (0.6) (EG2) 1-3, 1.6 (0.7) vs. 1.8 (0.7) vs. 1.7 (0.7) (CG) **(+) PPK (percent MET used):** 0-4, 1.0 (1.3) vs. 2.6 (1.2) vs. 2.8 (1.1) (EG) 0-4, 1.0 (1.2) vs. 1.4 (1.4) vs. 2.3 (1.2) (EG2) 0-4, 0.9 (1.2) vs. 1.1 (1.3) vs. 1.1 (1.2) (CG) **(+) PPK (MET techniques used):** 0-4, 2.5 (0.6) vs. 2.8 (0.6) vs. 2.8 (0.8) (EG) 0-4, 2.5 (0.7) vs. 2.5 (0.8) vs. 2.7 (0.8) (EG2) 0-4, 2.5 (0.7) vs. 2.5 (0.7) vs. 2.5 (0.6) (CG) **(+) M:** 1-5, 4.0 (0.6) vs. 4.4 (06) vs. 4.4 (0.6) (EG) 1-5, 3.9 (0.7) vs. 3.9 (0.6) vs. 4.2 (0.6) (EG2) 1-5, 3.8 (0.6) vs. 3.9 (0.7) vs. 3.8 (0.8) (CG)

Bellamy et al. 2014[c]	M: Field Notes, Focus Groups	**1 group pre vs. post:** **(+) M:** "Comparison of participant responses at baseline to the responses at follow up showed a general pattern of positive changes in participant perceptions of EBP." "In the second round of focus groups, participants described feeling more positive toward the concept of EBP and more motivated to use EBP in their agency." "Results from this study suggest that EBP training and partnership with researchers at schools of social work can be effective for motivating social work practitioners to adopt the EBP model [...]."
Bledsoe-Mansori et al. 2013[c]	S: Field Notes, Focus Groups	**1 group post-only, qualitative:** **(+) S:** "Overall this study indicates that the BEST training was generally acceptable to agency team members, though there are challenges that should be addressed to improve the acceptability of the training." "Overall, participants found the BEST training useful and beneficial [...]."
Bender et al. 2014	PDK: modified EBPPAS (α = .89) PPK: modified EBPPAS (α = .90) M: modified EBPPAS (α = .67 - .84) EBPPAS: Content, Criterion and Factorial Validity	**2 group pre vs. post (range 1-5, $\bar{x}$ (s):** **(+) PDK (familiarity):** 3.05 (.64) vs. 4.05 (.40), P < 0.001 (EG) 3.15 (.70) vs. 3.86 (.51), P < 0.001 (CG) **(+) PPK (engagement):** 3.01 (.63) vs. 3.32 (.59), P < 0.001 (EG) 3.12 (75) vs. 3.42 (.64), P < 0.01 (CG) **(+) M (attitude):** 3.74 (.34) vs. 3.82 (.44),P < 0.5 (EG) 3.72 (.38) vs. 3.81 (.42), P < 0.5 (CG) **(o) M (intentions):** 3.86 (.44) vs. 3.87 (.46) (EG) 3.86 (.52) vs. 3.90 (.55) (CG) **(-) M (feasibility):** 3.43 (.56) vs. 3.30 (.66), P < 0.5 (EG) 3.40 (.67) vs. 3.45 (.66) (CG)

Daniel et al. 2016	TDK: MCT (r = .84, content validity) PPK: Journal Entries M: Focus Group	**1 group pre vs. post (% (s)):** **(+) TDK:** 35 (0.13) vs. 70 (0.22), P = 0.01 (reported as P = .00) **(+) PPK:** "Quantitative analysis of the 30-day written survey found most (40%, n = 4) said they used music with 20% to 39% of their patients. 20% (n = 2) reported using music with 40% to 59% and 20% (n = 2) said they played music with 60% or more of their patients. Two staff members (20%) did not use the music protocol." **(+) M:** "Everyone in the sample strongly encouraged other social workers and chaplains to use the evidence-based music protocol with patients."
Dauenhauer et al. 2015	TDK: MCT (3 subscales, α = 0 - .533)	**1 group pre vs. post (range, $\bar{x}$ (s)):** **(+) TDK:** 0-20, 10.00 (2.25) vs. 13.31 (1.99), P = 0.001
Ducharme et al. 2015	PDK: OEQ PPK: LSQ (α = .92) S: OEQ	**1 group pre vs. post (range 1-10 , $\bar{x}$ (s)):** **(+) PDK of CBT:** 12 of 12 reported furthered knowledge, helpfulness: 8.2 (1.0) **(+) PDK of PT:** 11 of 12 reported furthered knowledge, helpfulness: 7.8 (1.2) **(+) PDK of IPT:** 6 of 12 reported furthered knowledge, helpfulness: 7.9 (1.4) **(range 1-100 , pooled $\bar{x}$ (range of s))** **(+) PPK:** 51.84 (14.05 - 26.05) vs. 73.52 (11.97-26.37), only 12 of 30 items reported **(+) S:** 8.6 (0.98)
Gromoske et al. 2015	PDK: EBPPAS-short (α = .91[d]) PPK: EBPPAS-short (α = .87[d]) M: EBPPAS-short (3 subscales, α = .79[d] - .87[d]) EBPPAS-short: Content, Criterion and Factorial Validity	**1 group pre vs. post vs. follow-up (range, $\tilde{x}$ (range)):** **(+) PDK (familiarity):** 1-50, 30 (31) vs. 40 (17) vs. 40 (29), P < 0.001 **(+) PPK (current use):** 1-30, 17 (20) vs. 19 (17), P < 0.01 **(+) M (intentions to use):** 1-30, 20 (19) vs. 22 (13) vs. 20 (17), P < 0.001 **(+) M (attitude):** 1-70, 52 (28) vs. 54 (22) vs. 54 (28), P < 0.001 **(o) M (feasibility):** 1-15, 10 (8) vs. 10 (8) vs. 10 (8)

Hagell et al. 2003	PPK: Focus Group M: Focus Group	**1 group post vs follow-up:** **(o) PPK:** "Eight out of the 20 respondents had looked at other research on care leavers as a result of listening to the tape. [...] Only one person had looked at any original reports or articles." **(o) M:** "Three months after listening to the tape, two-thirds of respondents in the follow-up interviews (14 out of 20) had not followed up any of the research or references cited on the tape."
Hohman et al. 2015	TPK: Vignettes (α = .76[d], ICC = .62 - .91) TPK: Vignettes (α = .87, ICC = .62 - .91)	**1 group pre vs. post (range, $\bar{x}$ (s)):** **(+) TPK:** 0–36, 11.04 (6.06) vs. 28.65 (4.32), P < 0.001 **(+) TPK:** 6–30, 7.08 (2.50) vs. 22.07 (4.94), P < 0.001
Holmes 2008	M: OEQ	**1 group post-only (range, $\bar{x}$ (s)):** **(+) M:** 1-10, 8 (unknown)
Kobak et al. 2016	TDK: MCT (α = .82) PDK: LSQ PPK: LSQ S: RS (α = .86) LSQ: construct validity	**1 group pre vs. post (range, $\bar{x}$ (s)):** **(+) TDK:** 24.4 (4.42) vs. 33.9 (5.11), P < 0.001 **(+) PDK, PPK (range, $\tilde{x}$):** 1-5, 4.4 **(+) S (technical aspects):** 0-100, 78.4 (20.44) **(+) S (clinical content):** 0-100, 84.4 (13.80)
Leathers et al. 2013	TDK: MCT (r = .56, discriminant validity) M: EBPAS (α = .75) S: LSQ (α = .75)	**2 groups pre vs. post vs. follow-up (range, $\bar{x}$ (s)):** **(+) TDK:** 0-160, 89 (11.16) vs. 94.78 (9.08), P < 0.5 vs. 94.11 (10.41), P < 0.1 (EG) 0-160, 95.00 (13.69) vs. 96.00 (9.79) vs. 95.22 (9.92) (CG) **(-/+) M:** 0-4, 3.14 (.43) vs. 2.98 (.36) vs. 3.17 (.32) (EG) 0-4, 3.01 (.31) vs. 2.88 (.23) vs. 2.87 (.30) (CG) **(+) S (online journals):** 0-4, 2.26 (0.77) (both groups)

Leathers et al. 2016	PPK: CQ	**2 group pre-repeated (used units of the intervention, $\bar{x}$ (s)):** **(+) PPK:** 0 vs. 4.7 (10.65) vs. 3.9 (7.3) vs. 1.67 (4.13) vs. .31 (.70) (EG) 0 vs. 0.83 (1.85) vs. 0.4 (1.39) vs. 0.52 (1.65) vs. 0 (CG) **(+) PPK:** "35.1 % of providers reported any use of the intervention materials at some point."
Lopez et al. 2011	PPK: LSQ (4 subscales, all $\alpha > .89$) M: EBPAS ($\alpha = .43 - .95$) M: LSQ ($\alpha = .62 - .81$) S: LSQ ($\alpha = .62 - .81$)	**1 group pre vs. post. vs. follow-up (range, $\bar{x}$ (s)):** **PPK:** "Difference scores were calculated comparing each scale score to the mean of the other three scales. A positive difference score indicates the scale was rated more positively than the other three models (on average)." **(+) PPK (behavioral):** 0.01 (0.44) vs 0.25 (0.54), $P < 0.5$ vs. 0.11 (0.46) **(+) PPK (family systems):** - 0.15 (0.65) vs. 0.03 (0.72) vs. 0.01 (0.52) **(+) PPK (cognitive):** 0.51 (0.48) vs. 0.37 (0.56) vs. 0.46 (0.40) **(+) PPK (psychodynamic):** -0.36 (0.67) vs. -0.65 (0.55), $P < 0.5$ vs. -0.58 (0.60) **(o) M (EBPs in general):** 1-4, 3.06 (0.43) vs. 3.09 (0.56) vs. 3.05 (0.53) **(+) M (BPT):** 0.14 (0.64) vs. 0.45 (0.70), $P < 0.01$ vs. 0.44 (0.66), $P < 0.05$ **(+) S:** 0.10 (1.0) vs. 0.70 (0.70), $P < 0.001$ vs. 0.67 (0.70), $P < 0.001$
Martin et al. 2013	TDK: MCT S: OEQ	**1 group pre vs. post ($\bar{x}$):** **(+) TDK:** "At the pre-test, most participants reported having limited or no understanding of EBP or EBHP." "At post-test the majority reported improvements in understanding EBP and EBHP concepts." (no further data reported) **(+) S:** 75%
Matthieu et al. 2016	PDK: EBPPAS M: EBPPAS EBPPAS: Content, Criterion and Factorial Validity	**1 group pre vs. post (range, $\bar{x}$ (s)):** **(+) PDK (Familiarity):** 1-50, 30.60 (6.54) vs. 40.92 (4.24), $P < 0.01$ **(+) M (Attitudes):** 1-70, 52.72 (6.07) vs. 57.57 (5.56), $P < 0.01$ **(+) M (Intentions):** 1-50, 36.41 (6.71) vs. 41.13 (6.11), $P < 0.01$

Parrish et al. 2011a	TDK: MCT PDK: EBPPAS (α = .91) PPK: EBPPAS (α = .86) M: EBPPAS (α = .63 - .86) EBPPAS: Content, Criterion and Factorial Validity	**1 group pre vs. post (range, $\bar{x}$ (s)):** **(+) TDK:** 57.05 (15.42) vs. 75.90 (11.31) P < 0.05 vs. 68.03 (12.49) **(+) PDK (Familiarity):** 0-50, 30.99 (6.86) vs. 41.48 (4.29), P < 0.05 vs. 40.33 (5.46) **(+) PPK (Engagement):** 28.60 (7.05) vs. 31.67 (6.96), P < 0.001 **(+) M (Attitudes):** 1-70, 48.61 (5.65) vs. 56.95 (6.29), P < 0.05 vs. 56.28 (7.05) **(+) M (Intentions):** 1-50, 32.57 (5.75) vs. 38.21 (4.47), P < 0.05 vs. 35.62 (6.32) **(+) M (Feasibility):** 1-35, 24.38 (3.06) vs. 27.03 (2.93), P < 0.05 vs. 26.05 (3.30)
Peterson et al. 2011	PDK: LSQ PPK: LSQ M: LSQ , OEQ	**1 group pre vs. post ($\bar{x}$ (s)):** **(+) PDK (Familiarity):** 3.1 (0.91), 3 vs. 4.2 (0.6), P < 0.001 **(+) PPK (ability to find relevant research):** 3.8 (.73), 4 vs. 4.2 (.56), P < 0.001 **(+) PPK (preparedness for EBPs use):** 3.1 (.84), 3 vs. 4.0 (.54), P < 0.001 **(+) M (OEQ):** "The EBP project also appeared to increased students' practice confidence […]." "I learned that I am a social worker; I went through all stages of GIM and applied research in my work. I can do it." "There were numerous comments about the project helping to clarify the relationship between research and practice and helping to develop research skills." "I learned how to incorporate evidence based practice into my internship. ... I also learned how important it is to research when I don't know the answer, and encouraging evaluation of [my] own practice." "I learned how important it is to consult credible research to better serve my clients."

Ronen 2005	PDK: OEQ PPK: OEQ M: OEQ S: LSQ (α = .82), OEQ	**1 group post-only (range 1- 7, pooled $\bar{x}$ (range of s)):** **(+) PDK:** "In a section asking for open-ended comments, the students reported that the project afforded them the most effective learning they had experienced that year and that they felt they had learned the most from this course and project." **(+) PPK:** "We learned effective techniques for intervention.' 'It has changed our way of intervening." **(+) M:** "[...] in this questionnaire, students commented that for the first time, they could really see how research affected the intervention and that the research seemed very beneficial." **(+) S:** 6.57 (0.00-0.93) "This was the most beneficial course we ever had."
Sacco et al. 2017	TDK: MCT (α = .48[d] knowledge questions and .96[d] case examples) TPK: Observation in roleplay (ICC = 82. - .96) PPK: LSQ (3 subscales, α = .95 - .96) PPK: LSQ (α = .91[d])	**1 group pre vs. post vs. follow-up (range, $\bar{x}$ (s)):** **(+) TDK:** 0-1, 0.58 (0.16) vs. 0.74 (0.13) vs. 0.70 (0.16) for knowledge questions 0-1, 0.91 (0.07) vs. 0.92 (0.13) vs. 0.92 (0.6) for case examples **(+) TPK:** 0-∞, 11.59 (4.04) vs. 15.50 (3.94) **(+) PPK (Screenings used):** 1-5, 2.80 (1.19) vs. 3.02 (1.10) vs. 3.07 (1.17) **(+) PPK (Brief Interventions):** 1-5, 2.28 (1.11) vs. 2.60 (1.18) vs. 2.45 (1.06) **(+) PPK (Referral for Treatment):** 1-5, 1.84 (1.03) vs. 2.13 (1.16) vs. 2.08 (0.98) **(+) PPK (Confidence):** 1-11, 6.15 (1.87) vs. 9.03 (1.11) vs. 8.88 (0.99)
Salcido 2008	PDK: EBPAS PPK: LSQ S: LSQ	**1 group pre vs. post (range 0-4, $\bar{x}$ (s)):** **(+) PDK (familiarity):** 1.68 (0.98) vs. 2.79 (0.74), P < 0.001 **(range 0-4, pooled $\bar{x}$ (range of s)):** **(+) PPK:** 1.25 (0.98-1.29) vs. 2.86 (0.76-1.10), P < 0.001 **(+) S:** 3.11 (0.70-0.98)
Salloum et al. 2013	S: Interview (κ = .86)	**1 group post-only:** **(+) S:** "[…] clinicians at one agency using the GTI Podcast series indicated that generally they found the Podcasts helpful for implementing a manualized treatment."

Smith et al. 2007	TPK: MCT PPK: LSQ M: LSQ	**1 group pre vs. post (range, pooled weighted $\bar{x}$)[e]:** **(o) TPK:** no significant change (p = 0.165), however, 3 out of 8 items significant at p < 0.1, but no values for the remaining 5 items provided **(+) PPK:** 1-5, 2.66 vs. 3.08 **(-) M:** 1-5, 2.80 vs. 2.70
Straussner et al. 2006	PDK: LSQ PPK: LSQ	**1 group pre vs. post (range, $\bar{x}$):** **(+) PDK, PPK:** 1-4, 2.71 vs. 2.95, P < 0.5
Tennille et al. 2016	PDK: EBPPAS-short (α = .89) PPK: EBPPAS-short (α = .89) M: EBPAS (α = .76), EBPPAS-short (α = .89) EBPPAS-short: Content, Criterion and Factorial Validity	**2 group pre vs. post vs. follow-up (range, $\bar{x}$):** **(+) PDK (EBP familiarity):** 1-50, 34.24 (1.32) vs. 39.07 (1.32) vs. 39.82 (1.33) (EG) 32.19 (1.26) vs. 32.26 (1.28) vs. 32.59 (1.27) (CG) **(+) PPK (current use):** 1-40, 19.76 (0.76) vs. 21.61 (0.82) vs. 22.51 (0.78) (EG) 17.45 (0.72) vs. 18.11 (0.79) vs. 18.67 (0.74) (CG) **(-) M (EBP attitude):** 1-70, 48.58 (0.92) vs. 44.73 (0.66) vs 44.02 (0.81) (EG) 47.38 (0.88) vs. 39.39 (0.63) vs. 39.28 (0.78) (CG) **(+) M (EBP feasibility):** 1-25, 20.08 (0.52) vs. 20.12 (0.56) .vs 20.82 (0.55) (EG) 19.50 (0.49) vs. 19.03 (0.54) vs. 19.03 (0.53) (CG) **(+) M (EBPs):** 1-4, 2.86 (0.07) vs. 3.08 (0.08) vs. 3.02 (0.08) (EG) 2.90 (0.07) vs. 2.76 (0.08) vs. 2.65 (0.07) (CG)

Webber et al. 2010	TDK: Concept Mapping PPK: LSQ, Focus Group S: Focus Group	**2 group pre vs. post (range, pooled $\bar{x}$):** **(o) TDK:** "Analysis of the students' concept maps revealed no conspicuous differences in the quality of student knowledge changes [...]." **(+) PPK:** 1-10, 4.87 (CG; n=12), 6.62 (EG; n=3) **(+) PPK:** "In particular the participants commented that the course gave them skills to evaluate research, which led to greater confidence in data interpretation and in interactions with other health professionals." "Being geared up to evaluate evidence based research gives me confidence." "With time, I definitely feel I will be able to critique research more confidently." **(+) S:** "The course has been excellent at introducing the concepts and the online seminars were of particular use in developing understanding." "The course has enabled me to gain greater confidence in understanding and appraising research." "Met expectations over and above." "I feel the course delivered what it promised over an appropriate time frame."
Wong 2017	TPK: Assignment PDK: OEQ (κ = .87) PPK: OEQ (κ = .87) M: OEQ (κ = .87) S: LSQ	**1 group post-only (range, pooled $\bar{x}$ (range of s)):** **(+) TPK:** 1-4, 3.31 (0.4-0.9) **(+) PDK:** "I feel the assignment taught me really good, hands-on things, for example, what databases to search for evidence-based interventions." "I learned the criteria for various levels of evidence [...]" **(+) PPK:** "I was able to research EBP and use scholarly articles to see how various models of evidence-based interventions were applied to diverse populations" "I really like being able . . . to evaluate while I read. Now, I can evaluate studies better." **(+) M:** "It helped me realize the importance of staying current on EBP." "One needs to learn how to read, understand, and evaluate empirical studies." **(+) S:** 1-4, 3.4 (0.8-1.2)

Note. EBP = Process of Evidence-based Practice, (+) = 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 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, $\bar{x}$ = mean, s = standard deviation, $\tilde{x}$ = median, EG = Experimental Group, CG = Control Group

a if necessary, resp. if several constructs that were measured in the study apply to one coded DV
b delayed intervention (after T2)
c articles refer to the same study
d mean value, computed from n values (n = points of measurement with reference to the study design)
e the study reports a positive effect on 'Motivation', however, due to our coding scheme we allocated 7 out 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". Afterwards, three of the remaining four items were allocated to the construct 'Motivation'. An example of an affected item would be "Original research is confusing".