

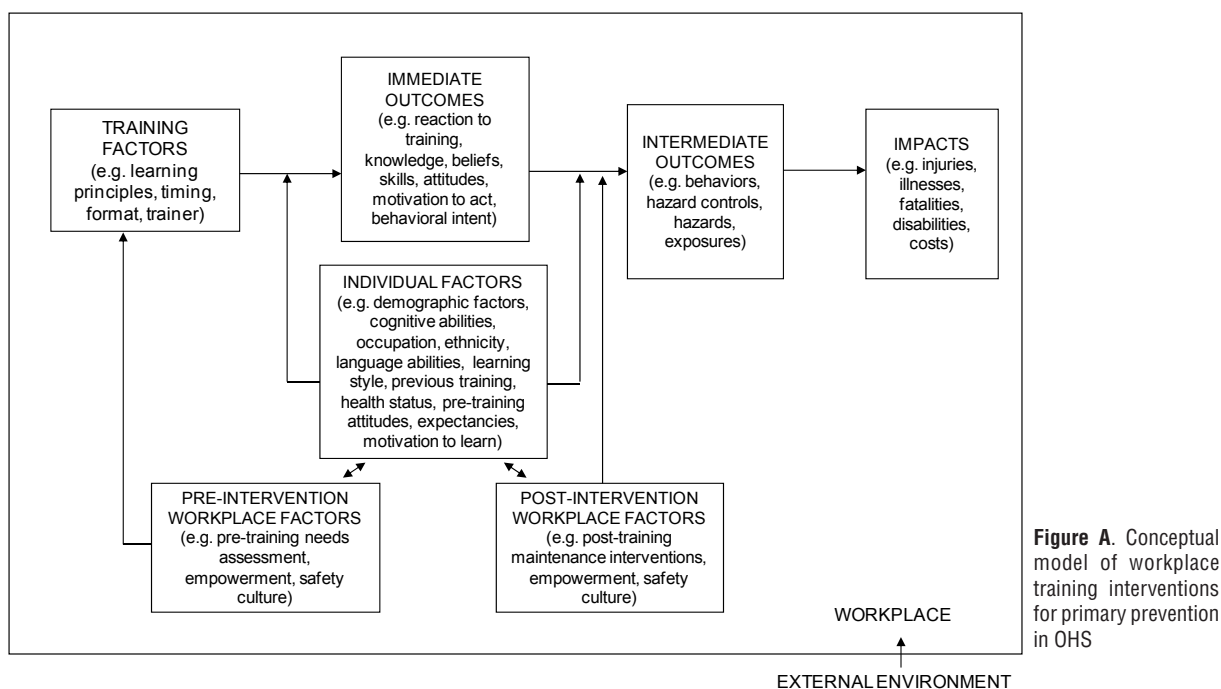
A systematic review of the effectiveness of occupational health and safety training ¹

by Lynda S Robson, PhD,² Carol M Stephenson, PhD, Paul A Schulte, PhD, Benjamin C Amick III, PhD, Emma L Irvin, BA, Donald E Eggerth, PhD, Stella Chan, MSc, Amber R Bielecky, MSc, Anna M Wang, BScN, Terri L Heidotting, EdD, Robert H Peters, MSc, Judith A Clarke, MA, Kimberley Cullen, MSc, Cathy J Rotunda, EdD, Paula L Grubb, PhD

1 Appendix – Supplementary tables A–K

2 Corresponding author Institute for Work & Health, 481 University Ave, Suite 800, Toronto, ON M5G 2E9, Canada. [E-mail: lrobson@iwh.on.ca]

Appendix: Figure A and Tables A–K

Table A. Terms for literature search ^a

Term Category	Exact Search Terms ^b		
Work-related	employment/ occupations	work/ worker\$.mp./	workplace\$.mp.
Education and training intervention	blended training.mp. education/ educational measurement/ educational status/ e-training.mp. extra training.mp.	health education/ Health Education/mt [Methods] health knowledge, attitudes, practice/ inservice training/ Inservice Training/ec, og, mt intervention?.mp.	on demand training.mp. pre training.mp. training.mp.
OHS outcomes and factors affecting effectiveness	absenteeism/ accidents/ accidents, occupational/ accident prevention/ barriers.mp. behavior?al change.mp. claim\$.mp. "Costs and Cost Analysis"/ cumulative trauma disorder/ economic evaluation.mp. Evaluation Studies/ facilitators.mp.	feedback procedures.mp. feedback/ hazardous substances/ hazardous waste/ Health Knowledge, Attitudes, Practice/ health protection.mp. KNOWLEDGE/ medical care.mp. occupational diseases/ occupational exposure/ occupational health/ Occupational Health Services/	performance indicators.mp. presenteeism.mp. primary prevention/ protective factors.mp. practice.mp. reinforcement.mp. return on investment.mp. Risk Factors/ safety/ safety culture.mp. workers' compensation/ "wounds and injuries"/
Between-group evaluation design	between groups.mp. comparison.mp.		random\$.mp.

^a The search was limited to humans and (English or French) and yr= "1996 – 2007". Terms were also used to *exclude* publications: health promotion, diet, exercise, smoking, weight loss and addiction.

^b The terms are presented using Medline syntax. This was adapted for other databases. Search terms were combined using the following Boolean logic: terms within the same category were combined using "OR" and the terms in different categories were combined using "AND."

Table B. Relevance assessment questions, stage 1. [OHS=occupational health and safety.]

Question	Exclusionary response
1 Is the study examining a worker population?	No
2 Is the study concerned with any of the following? An intervention study (with pre- and post- measures) assessing the effectiveness of an OHS education/training program Factors that may facilitate or inhibit the effectiveness of OHS education/training programs A novel approach to provide OHS education/training programs Specialized techniques/methods (eg, computer-based training) that have been used to provide OHS education/training programs Factors that affect compliance with OHS education/training programs	No
3 Does the study present information that is best described as “conjecture” or “testimonials” with no supporting evidence?	Yes
4 Does the study focus on workers’ current state of knowledge regarding an OHS issue, which simply identifies that there is a further need for education/training on this issue?	Yes
5 Is the study published in English or French?	No
6 Is the date of publication between 1996 and 2007?	No

Table C. Effect of training on knowledge (relative to a no-training control). [ANCOVA=analysis of covariance; *d*=standardized mean difference; NR=not reported; SE=standard error.]

Authors	Intervention ^a (level of engagement ^b number of training sessions; hazard type ^c)	Method limitations score ^d	N ^e	Time of follow-up	Outcome (method of measurement) ^f	Results, as reported in original publication		Effect size ^g	
						Direction of effect ^h	Statistical significance ^h (between group)	<i>d</i>	SE
Gray et al (63)	Multi-component lift training for nurses (H;5;E)	8	NR	0	Knowledge (Q)	+	P<0.001	..	
Greene et al (34)	Multi-component office ergonomics (H;2;E)	4	83	2 weeks	Knowledge and beliefs (Q)	+	P<0.01 for inter- action of baseline knowledge and group in ANCOVA	1.45	0.25
Harrington & Walker (45)	Computer-based home office ergonomics training (L;1;E)	4	50	0	Knowledge (Q)	+	P<0.0001 [#]	3.58	0.46
Harrington & Walker (56)	1) Computer-based fire safety training at life care facility (M;2;P) 2) Instructor-led fire safety training (L;2;P)	5	1) 46 2) 45	0	i) Knowledge on hazards (Q) ii) Knowledge on safety behaviors (Q)	1(i) + [#] 1(ii) + [#] 2(i) + [#] 2(ii) + [#]	ANCOVA 1(i) P<0.05 1(ii) P<0.05 2(i) P<0.05 2(ii) P<0.05	1(i) 1.30 1(ii) 1.06 2(i) 0.94 2(ii) 0.83	0.34 0.31 0.31 0.32
Rizzo et al (47)	1) Instructor-directed computer ergonomics (L;1;E) 2) Video and pamphlets; computer ergonomics (L;1;E)	5	1) 45 2) 39	15 months	Knowledge (Q)	1) + 2) +	ANCOVA 1) P<0.05 2) P<0.05	1) 1.38 2) 0.78	0.34 0.34

^a More details on the interventions are in Table 4.

^b Level of learner engagement in the intervention, as described in methods: low (L), medium (M), or high (H).

^c Hazard type: biological (B), chemical (C), ergonomic (E), physical (P), safety (S).

^d Methodological limitations score as described in Methods, with range 0 (no limitations) to 8 (most limitations).

^e Total number of subjects in the intervention and control groups in analysis. Subjects are people unless indicated otherwise.

^f Method used to measure outcome: administrative records (A), clinical exam (C), observations (O), physical property measurement (P), questionnaire/diary (Q), voluntary registry (R).

^g Direction of effect, as reported by original authors for between-group results. When not reported by the original authors, the determination was made by the reviewers (indicated by #), guided by the authors’ approach to analysis. A positive value indicates training more effective than control.

^h Statistical significance, as reported by original authors for a between-group test. When not reported by the original authors, it was determined by reviewers (indicated by #) when baseline similarity of groups with respect to the outcome could be established (using criterion of $p>0.05$). Tests were conducted on post-intervention data only, unless indicated otherwise.

ⁱ *d* was calculated by reviewers from post-intervention data when baseline similarity of the groups with respect to the outcome could be established (using criterion of $P>0.05$). A positive value indicates training more effective than control.

Table D. Effect of training on attitudes and beliefs (relative to a no-training control). [ANCOVA=analysis of covariance; d= standardized mean difference; NS=not statistically significant; NC=not calculated by reviewers because of baseline dissimilarity of groups with respect to outcome measures; SE=standard error]

Authors	Intervention ^a (level of engagement ^b number of training sessions; hazard type ^c)	Method limitations score ^d	N ^e	Time of follow-up	Outcome (method of measurement) ^f	Results, as reported in original publication		Effect size ⁱ	
						Direction of effect ^g	Statistical significance ^h (between group)	d	SE
Greene et al (34)	Multi-component office ergonomics (H;2;E)	4	83	2 weeks	i) Self-efficacy (Q) ii) Outcome expectations (Q)	i)+ ii)+	ANCOVA: i) P<0.01 for interaction of baseline self-effi- cacy and group ii) P=0.00	i) 0.82 ii) 0.87	0.23 0.23
Harrington & Walker (45)	Computer-based home office ergonomics training (L;1;E)	4	50	0	Attitudes (Q)	+	P=0.0001 [#]	1.41	0.32
Harrington & Walker (56)	1) Computer-based fire safety training at life care facility (M;2;P) 2) Instructor-led fire safety training (L;2;P)	5	1) 46 2) 45	0	ii) Attitudes to hazards (Q) ii) Attitudes to safety behaviors (Q)	1(i) - 1(ii) + 2(i) - 2(ii) -	ANCOVA on groups 1,2 and control 1(i) NS 1(ii) NS 2(i) NS 2(ii) NS	1(i) NC 1(ii) 0.21 2(i) -0.03 2(ii) -0.13	0.29 0.29 0.31

^a More details on the interventions are in Table 4.^b Level of learner engagement in the intervention, as described in methods: low (L), medium (M), or high (H).^c Hazard type: biological (B), chemical (C), ergonomic (E), physical (P), safety (S).^d Methodological limitations score as described in Methods, with range 0 (no limitations) to 8 (most limitations).^e Total number of subjects in the intervention and control groups in analysis. Subjects are people unless indicated otherwise.^f Method used to measure outcome: administrative records (A), clinical exam (C), observations (O), physical property measurement (P), questionnaire/diary (Q), voluntary registry (R).^g Direction of effect, as reported by original authors for between-group results. When not reported by the original authors, the determination was made by the reviewers (indicated by #), guided by the authors' approach to analysis. A positive value indicates training more effective than control.^h Statistical significance, as reported by original authors for a between-group test. When not reported by the original authors, it was determined by reviewers (indicated by #) when baseline similarity of groups with respect to the outcome could be established (using criterion of p>0.05). Tests were conducted on post-intervention data only, unless indicated otherwise.ⁱ d was calculated by reviewers from post-intervention data when baseline similarity of the groups with respect to the outcome could be established (using criterion of P>0.05). A positive value indicates training more effective than control.**Table E.** Effect of training on behaviors (relative to a no-training control). [ANCOVA=analysis of covariance; d=standardized mean difference; NC=not calculated by reviewers because of baseline dissimilarity of groups with respect to outcome measures; NS=not statistically significant; PPE=personal protective equipment; PS=psychosocial; RM ANOVA=repeated measures analysis of variance; SE=standard error; UP=universal precautions.]

Authors	Intervention ^a (level of engagement ^b number of training sessions; hazard type ^c)	Method limita- tions score ^d	N ^e	Time of follow-up	Outcome (method of measurement) ^f	Results, as reported in original publication		Effect size ⁱ	
						Direction of effect ^g	Statistical significance ^h (between group)	d	SE
Arnetz & Arnetz (50)	Group discussion of violent incidents (both groups with violent incident forms) (M; multiple; S)	5	686	0	Exposures to vio- - lent incidents (Q)		P=0.03	-0.20	0.08
Bohr (53, 54)	1) Multi-component office ergonom- ics (H;1;E) 2) Lecture (L;1;E)	6	1) 103 2) 104	12 months	i) Workstation hazards (O) ii) Postural behaviors (O)	1(i) - # 1(ii) - # 2(i) + # 2(ii) + #	P≥0.05 (RM ANOVA on all 3 groups)	..	
Brisson et al (31)	Multi-component office ergonomics (H;2;E) for each of: 1) <40 years 2) ≥ 40 years	4	1) 207 2) 420	6 months	i) 3 postural behaviors (O) ii) 10 workstation hazards (O)	All # 1(i) NC, +, + 1(ii) +, NC, NC, NC, NC, +, -, NC, +, + 2(i) +, NC, + 2(ii) NC, +, +, +, +, +, NC, +, +, +	All # 1i) P=NC, 0.00, 0.52 1ii) P=0.02, NC, NC, NC, NC, 0.02, 0.15, NC, 0.10, 0.04 2(i) P=0.99, nc, 0.00 2(ii) P=NC, 0.00, 0.07, 0.15, 0.01, 0.00, NC, 0.01, 0.22, 0.00	1(i) NC, 0.49, 0.11 1(ii) 0.34, NC, NC, NC, NC, 0.35, -0.22, NC, 0.29, 0.30 2(i) 0.01, NC, 0.35 2(ii) NC, 0.38, 0.19, 0.15, 0.26, 0.31, NC, 0.30, 0.14, 0.30	1(i) 0.14 1(ii) 0.14 2(i) 0.10 2(ii) 0.10

Table E. Continued

Authors	Intervention ^a (level of engagement ^b number of training sessions; hazard type ^c)	Method limita- tions score ^d	N ^e	Time of follow-up	Outcome (method of measurement) ^f	Results, as reported in original publication		Effect size ^g	
						Direction of effect ^g	Statistical significance ^h (between group)	<i>d</i>	SE
Eklöf et al (40); Eklöf & Hagberg (46)	Multi-component office ergonomics including PS factors (H;1;E) Three study arms with different targets for feedback: 1) individual 2) supervisors 3) group	3	1) 18 2) 18 3) 18 (unit: work group)	6 months	i) % ergonomic modifications (Q) ii) Average num- ber ergonomic modifications (Q)	1(i) + 2(ii) + 3(i) + 1(iii) + # 2(ii) + # 3(ii) + #	1(i) P=0.02 2(ii) P=0.02 3(i) P=0.06 1,2,3 (ii) P=0.24 in overall test All based on change data.	1(i) 1.09 2(i) 1.71 3(i) +1.98 1(ii) 0.95 2(ii) 1.35 3(ii) 2.36	0.51 0.56 0.59 0.50 0.53 0.63
Greene et al (34)	Multi-component office ergonomics (H;2;E)	4	87	2 weeks	Postural expo- sures (O)	+	P<0.01 for interaction of postural exposures and group in ANCOVA	1.16	0.23
Held et al (33)	Formalized education program on skin care (H;3;C)	0	287	5 months	6 wet work behaviors (Q)	i) + ii) + iii) .. iv) .. v) + vi) + #	Using change data: i) P=0.02 ii) P<0.001 iii) P=0.80 iv) P=0.54 v) P=0.06 vi) P= 0.11	i) 0.51* ii) 0.83* iii) 0.03* iv) 0.17* v) 0.52* vi) 0.32*	0.12 0.12 0.12 0.12 0.12 0.12
Jensen et al (51)	Train-the-trainer program in patient lifting techniques (H;2;E)	6	~64	9 months	i) Physical exer- tion (Q) ii) Handling be- havior (Q) iii) Range of mo- tion (O)	i) - ii) .. iii) ..	i) NS ii) NS iii) NS	i) -0.17 ii) .. iii) ..	0.25
Rasmussen et al (49)	Multi-component farm safety (H;2;S)	1	~171 farms	6 months	i) Active safety behaviors (Q) ii) PPE use (Q)	i) + ii) +	Using change data: i) P=0.035 ii) P=0.005	..	
Rizzo et al (47)	1) Instructor-directed computer ergonomics (L;1;E) 2) Video and pamphlets; computer ergonomics (L;1;E)	6	1) 45 2) 39	15 months	Work habits (Q)	1) + 2) +	ANCOVA: 1) P<0.05 2) P<0.05	1) NC 2) 1.32	0.36
Wright et al (52)	Computer-assisted with problem- solving; UP behaviors (M;1;B)	3	60	2 months	UP behaviors (O)	+	Using change data: P=0.0004	1.25	0.28

^a More details on the interventions are in Table 4.

^b Level of learner engagement in the intervention, as described in methods: low (L), medium (M), or high (H).

^c Hazard type: biological (B), chemical (C), ergonomic (E), physical (P), safety (S).

^d Methodological limitations score as described in Methods, with range 0 (no limitations) to 8 (most limitations).

^e Total number of subjects in the intervention and control groups in analysis. Subjects are people unless indicated otherwise.

^f Method used to measure outcome: administrative records (A), clinical exam (C), observations (O), physical property measurement (P), questionnaire/diary (Q), voluntary registry (R).

^g Direction of effect, as reported by original authors for between-group results. When not reported by the original authors, the determination was made by the reviewers (indicated by #), guided by the authors' approach to analysis. A positive value indicates training more effective than control.

^h Statistical significance, as reported by original authors for a between-group test. When not reported by the original authors, it was determined by reviewers (indicated by #) when baseline similarity of groups with respect to the outcome could be established (using criterion of P>0.05). Tests were conducted on post-intervention data only, unless indicated otherwise.

ⁱ *d* was calculated by reviewers from post-intervention data when baseline similarity of the groups with respect to the outcome could be established (using criterion of P>0.05). A positive value indicates training more effective than control. In some cases, data were estimated from a Figure (indicated by *).

Table F. Effect of training on health (relative to a no-training control). [ANCOVA=analysis of covariance; *d*=standardized mean difference; MSK=musculoskeletal; NS=not statistically significant; PS=psychosocial; SE=standard error; UE=upper extremity; US=upper spine.]

Authors	Intervention ^a (level of engagement ^b number of training sessions; hazard type ^c)	Method limita- tions score ^d	N ^e	Time of follow-up	Outcome (method of measurement) ^f	Results, as reported in original publication		Effect size ⁱ	
						Direction of effect ^g	Statistical significance ^h (between group)	<i>d</i>	SE
Banco et al (55)	Training on old cutter use with retail employees (H;1;S)	4	670	12 months	Cutter-related injury rate (A)	-	p = 0.8 [#]	Rate ratio: 0.90	0.42
Bohr (53, 54)	1) Multi-component office ergonomics (H;1;E) 2) Lecture (L;1;E)	6	1) 85 2) 86	12 months	Upper-body MSK symptom score (Q)	1) + [#] 2) + [#]	1,2) P<0.01 (RM ANOVA on all 3 groups)	..	
Brisson et al (31)	Multi-component office ergonomics (H;2;E) for each of: 1) <40 years 2) ≥ 40 years	4	1) 162 2) 328	6 months	MSK disorder prevalence (C)	1) + 2) - [#]	1) P=0.1 [#] 2) P=0.4 [#]	1) 0.33 2) -0.11	0.16 0.11
Duffy & Hazlett (32)	1) Information on voice use for teaching trainees (L;1;E) 2) Practical training on voice use for teaching trainees (H;2;E)	6	1) 35 2) 43	2 months	Voice quality score (P)	1) + 2) +	P=0.18 (RM ANOVA on all 3 groups)	1) 0.04 * 2) 0.30 *	0.36 0.31
Eklöf et al (40); Eklöf & Hagberg (46)	Multi-component office ergonomics including PS factors (H;1;E) Three study arms with different targets for feedback: 1) individual 2) supervisors 3) group	3	1) 18 2) 18 3) 18 (unit: work group)	6 months	MSK or eye symp- tom prevalence (Q)	1) - [#] 2) - [#] 3) - [#]	1,2,3) P=0.90 in overall test of change data	1) -0.13 2) -1.34 3) -0.37	0.47 0.53 0.48
Greene et al (34)	Multi-component office ergonomics (H;2;E)	2	82	2 weeks	MSK symptom score (Q): i) UE intensity ii) UE frequency iii) UE duration iv) US intensity v) US frequency vi) US duration	i) - ii) + iii) - iv) + v) + vi) +	i) P=0.6 ii) P=0.9 iii) P=0.7 iv) P=0.8 v) P=0.4 vi) P=0.7	i) -0.12 ii) 0.03 iii) -0.15 iv) 0.15 v) 0.27 vi) 0.37	0.22 0.22 0.22 0.22 0.22 0.22
Held et al (33)	Formalized education program on skin care (H;3;C)	0	287	5 months	Skin symptom severity (C)	+	P=0.0002 (based on change data)	0.05*	0.12
Jensen et al (51)	Train-the-trainer program in patient lifting techniques (H;2;E)	6	114	15 months	Low back pain (Q): i) past year ii) past 3 months	i) - ii) -	Using change data: i) P=0.10 ii) P=0.16	i) 0.04 ii) 0.00	0.19 0.19
Rasmussen et al (49)	Multi-component farm safety (H;2;S)	1	~930	6 months	Farm-related injury rates; all injuries (Q)	+	P=NS (Poisson regression model)	Rate ratio: 0.91	0.22
van Poppel et al (58)	Lifting ergonomics for material han- dlers (H;3;E)	5	282	6 months	Low-back pain past month prevalence (Q)	-	P=0.97	-0.004	0.12

^a More details on the interventions are in Table 4.^b Level of learner engagement in the intervention, as described in methods: low (L), medium (M), or high (H).^c Hazard type: biological (B), chemical (C), ergonomic (E), physical (P), safety (S).^d Methodological limitations score as described in Methods, with range 0 (no limitations) to 8 (most limitations).^e Total number of subjects in the intervention and control groups in analysis. Subjects are people unless indicated otherwise.^f Method used to measure outcome: administrative records (A), clinical exam (C), observations (O), physical property measurement (P), questionnaire/diary (Q), voluntary registry (R).^g Direction of effect, as reported by original authors for between-group results. When not reported by the original authors, the determination was made by the reviewers (indicated by #), guided by the authors' approach to analysis. A positive value indicates training more effective than control.^h Statistical significance, as reported by original authors for a between-group test. When not reported by the original authors, it was determined by reviewers (indicated by #) when baseline similarity of groups with respect to the outcome could be established (using criterion of P>0.05). Tests were conducted on post-intervention data only, unless indicated otherwise.ⁱ *d* was calculated by reviewers from post-intervention data when baseline similarity of the groups with respect to the outcome could be established (using criterion of P>0.05). A positive value indicates training more effective than control. In some cases, data were estimated from a Figure (indicated by *).

Table G. Relative effectiveness of differing levels of engagement on outcomes. [ANCOVA=analysis of covariance; *d*=standardized mean difference; HPD=hearing protection device; I=intervention; NC=not calculated by reviewer because baseline dissimilarity of groups with respect to outcome measure; NS=not statistically significant; PPE=personal protective equipment; Q&A=question & answer session; RM ANOVA=repeated measures analysis of variance; SE=standard error; UE MSK=upper extremity musculoskeletal]

Authors	Intervention ^a (level of engagement ^b number of training sessions; hazard type ^c)	Method limitations score ^d	N ^e	Time of follow-up	Outcome (method of measurement) ^f	Results, as reported in original publication		Effect size ⁱ	
						Direction of effect ^g	Statistical significance ^h (between group)	<i>d</i>	SE
Bohr (53, 54)	I1) Participatory education (hands-on demo, problem solving, application to work area) (H;1;E) I2) Traditional education (lecture, informational handout, Q&A session) (L;1;E)	6	101	12 months	i) Workstation hazards (O) ii) Postural behaviors (O) iii) UE MSK symptoms (Q)	I1 ver- sus I2: i) – # ii) – # iii) + #	I1 versus I2: RM ANOVA: i) NS ii) NS iii) NS	I1 versus I2: i) .. ii) .. iii) ..	
Duffy & Hazlett (32)	I1) Direct voice care training (vocal- ization, posture, respiration, release of tension in vocal apparatus, reso- nance, and voice projection) (H;2;E) I2) Indirect voice care training (infor- mation on voice production, factors associated with healthy voice) (L;1;E)	6	32	2 months	Voice quality (Dysphonia Severity Index) (P)	I1 ver- sus I2: i) +	1, 2, control: RM ANOVA: P=0.18	I1 versus I2: 0.28	0.37
Harrington & Walker (56)	I1) Computer-based instruction on fire safety; screens contained narra- tion, interaction, animation or video; some with questions and interactive games (M;2;S) I2) Lectures & printed materials (L;2;S)	5	45	0	Fire safety outcomes (Q): i) knowledge on hazards ii) knowledge on safety behaviors iii) attitudes on hazards iv) attitudes on safety behaviors	I1 ver- sus I2: i) + # ii) + # iii) + # iv) + #	I1 versus I2: i) P=NS ii) P=NS 1,2, control: ANCOVA: iii) P=NS iv) P=NS	I1 versus I2: i) 0.41 ii) 0.26 iii) 0.56 iv) 0.21	0.31 0.30 0.31 0.30
Hong et al (62)	I1) Computer w/ tailored feedback incl. hearing test result; HPD practice (H;1;P) I2) Commercial video w/ hearing test result (M;1;P)	i) 2 ii) 3	403	i) 0 wks ii) 12 months	i) Intention to use HPDs (Q) ii) Use of HPDs (Q)	i) + ii) +	I1 versus I2: RM ANOVA: i) P=0.001 ii) NS	I1 versus I2: i) 0.13 ii) 0.03	0.10 0.10
Löffler et al (42)	I1) Lecture, problem-solving and practice regarding skin care (M;7;C) I2) Informational paper on skin care (L;1;C)	3	325	2.5 years after 1st session	Dermatitis (C)	+	I1 versus I2: Multiple logis- tic regression: P=0.0001	I1 versus I2: 0.60	0.15
Lusk et al (65, 66)	I1) Tailored: Computer-based train- ing tailored to worker's self-reported practice. Used factual, cognitive approaches, demonstration, di- rected practice, vicarious experi- ence, persuasion and role-modeling techniques. Presented in interactive format, with feedback (H;1;P) I2) Non-tailored: As above, but deliv- ered to all participants in a uniform manner (H;1;P) I3) Video (L;1;P)	4	879	6–18 months	Use of hearing protec- tion (Q)	I1 ver- sus I3: + I2 ver- sus I3: -	RM ANOVA: 1 versus 3: P=0.49 I2 versus 3: P=0.18	I1 versus I3: NC I2 versus I3: 0.02	0.07
Perry & Layde (57)	I1) Education intervention (lecture, slides, presentation by respected area farmer, demonstration and opportunity for hands-on practice) (H;1;C) I2) Standard re-certification meeting for pesticide applicators (L;1;C)	i) 2 ii) 1 iii) 1 iv) 1 v) 1 vi) 1	385	6 months	i) Safety knowledge (Q) ii) PPE use other than gloves (Q) iii) Full PPE compli- ance (Q) iv) Dermal exposure (Q) v) Safety intentions (Q) vi) Risk perceptions (Q)	I1 ver- sus I2: i) + ii) + iii) + iv) + v) + vi) +	I1 versus I2: i) P<0.05 ii) P<0.05 iii) NS iv) NS v) P<0.05 vi) P<0.05	I1 versus I2: i) .. ii) +0.23 iii) +0.05 iv) +0.08 v) .. vi) ..	0.10 0.10 0.04

^a More details on interventions in Table 4. Level of engagement, as described in Methods: low (L), medium (M) or high (H).

^b Level of learner engagement, as described in Methods: low (L), medium (M) or high (H).

^c Hazard type: biological (B), chemical (C), ergonomic (E), physical (P), safety (S).

^d Methodological limitations score, as described in Methods, with range 0 (no limitations) to 8 (most limitations).

^e Total number of subjects in the intervention and control groups in analysis. Subjects are people unless indicated otherwise.

^f Method used to measure outcome: clinical exam (C), observations (O), physical property measurement (P), questionnaire/diary (Q).

^g Direction of effect: “+” indicates the higher level of engagement training was more effective than lower, as reported by the original authors for between-group results, when available. When not reported, the determination was made by the reviewers (indicated by #), guided by the authors’ approach to analysis.

^h Statistical significance: As reported by original authors for a between-group test, when available. When not reported, it was calculated by reviewers (indicated by #) when baseline similarity could be established (using criterion P>0.5). Tests were conducted on post-intervention data only, unless indicated otherwise.

ⁱ Effect size was calculated from post-intervention data when baseline similarity of the groups with respect to the outcome could be established, using criterion of P>0.5. “+” indicates training with the higher level of engagement was more effective than low.

Table H. Algorithm applied to higher versus lower engagement training evidence to determine its strength. [*d*=standardized mean difference; IQR=interquartile range.]

Strength of evidence	Methodological quality ^a	Minimum quantity	Consistency of effects ^b	Effect size criterion ^c
Strong	Good	≥2 studies	IQR (or range) of <i>d</i> does not include zero	Sufficient: Knowledge = 0.25 Attitudes & Beliefs = 0.12 Behaviors = 0.10 Health = 0.04
	Good or Fair	≥5 studies	IQR (or range) of <i>d</i> does not include zero	Sufficient: Knowledge = 0.25 Attitudes & Beliefs = 0.12 Behaviors = 0.10 Health = 0.04
	Meet methodological quality, quantity and consistency criteria for sufficient but not strong evidence			Large: Knowledge = 0.38 Attitudes & Beliefs = 0.25 Behaviors = 0.20 Health = 0.08
Sufficient	Good	Not applicable	IQR (or range) of <i>d</i> does not include zero	Sufficient: Knowledge = 0.25 Attitudes & Beliefs = 0.12 Behaviors = 0.10 Health = 0.04
	Good or Fair	≥3 studies	IQR (or range) of <i>d</i> does not include zero	Sufficient: Knowledge = 0.25 Attitudes & Beliefs = 0.12 Behaviors = 0.10 Health = 0.04
Insufficient	The criterion in any one of the four domains not met			Health = 0.04

^a Methodological quality categories for studies: Good (0–1 limitations score), Fair (3–4), and Limited (≥5).

^b Interquartile range of *d* was determined when there were five or more effect sizes in the body of evidence; otherwise the full range was used.

^c Criteria for sufficient and large effect sizes were defined by the research team. These were operationalized by requiring the median effect size for a body of evidence to be equal to or greater than the criterion.

Table I. Final bodies of evidence on the relative effectiveness of higher versus lower engagement training, by outcome. [*d*=standardized mean difference; FB=feedback; SE=standard error]

Authors; intervention ^a (specific levels of learner engagement ^b ; # of sessions)	Methodological quality ^c	Calculated effect sizes							
		Knowledge		Attitudes		Behaviors		Health	
		<i>d</i>	SE	<i>d</i>	SE	<i>d</i>	SE	<i>d</i>	SE
Hong et al (62); hearing protection (H versus M;1)	Fair	..	0.13	0.10	0.03	0.10	..	..	..
Löffler et al (42); dermatitis prevention (M versus L;7 versus 1)	Fair	..	..	..	..	..	..	0.60	0.15
Lusk et al (65, 66); hearing protection (H versus L;1)	Fair	..	..	..	0.02	0.07	..	..	..
Perry & Layde (57); safe pesticide use (H versus L;1)	Fair	+ ^e	..	..	..	..	..	..	..
Perry & Layde (57); safe pesticide use (H versus L;1)	Good	..	..	..	ii–iii) 0.14 *	..	..	..	..
					iv) 0.08	0.04			

^a More details of interventions are in Table 4.

^b Specific levels of learner engagement contrasted in the study, determined as described in Methods: low (L), medium (M), or high (H).

^c Refers to the assessed methodological quality of the specific outcome data included in the table: Good, 0–1 methodological limitations score; Fair, 2–4.

^d *d* values and SE from Table G were carried over to Table I when the methodological quality of the data was fair or good. When a study reported multiple measures of a similar concept, only their median reported here (indicated by *). The standard error for each of the contributing individual measures can be found in Table G. A positive value for *d* indicates the higher engagement training is more effective than the lower engagement training. Numbered bullets preceding individual effect sizes correspond to those used in Table G.

^e Effect size not calculable, but direction is indicated.

Table J. Summary of methodological quality assessments of studies. [K=knowledge; A=attitudes and beliefs; B=behaviors; H=health; CSG=comparability of study groups; II=intervention implementation; OA=outcome assessment; SA=statistical analysis.]

Authors	Initial study sample size ^a	Outcomes assessed	Methodological assessments: four domains ^b				Methodological limitations score used in evidence synthesis ^d
			CSG	II	OA	SA	
Arnetz & Arnetz (50)	*1500	B	P	P	N	P	5
Banco et al (55)	*900	H	P	P	P	P	4
Bohr (53, 54)	154	B, H	N	N	Y	N	6
Brisson et al (31)	*658	B, H	P	P	Y	N	4
Duffy & Hazlett (32)	55	H	N	N	Y	N	6
Eklöf et al (40); Eklöf & Hagberg (46)	396	B, H	P	P	P	Y	3
Gray et al (63)	*250	K	N	N	N	N	8
Greene et al (34)	87	K, A, B, H	P	P	P	P	4
Harrington & Walker (45)	102	K, A	P	N	P	Y	4
Harrington & Walker (56)	141	K, A	N	P	P	P	5
Held et al (33)	375	B, H	Y	Y	Y	Y	0
Hickman & Geller (44)	15	B	Y	P	Y	Y	1
Hong et al (62)	612	A, B	P	Y	Y	P	2
Jensen et al (51)	210	B, H	P	P	N	N	6
Löffler et al (42)	521	H	P	P	Y	P	3
Lusk et al (65, 66)	2219	B	P	P	P	P	4
Perry & Layde (57)	400	K, B	Y	P	Y	Y	1
Rasmussen et al (49)	990	B, H	Y	Y	P	Y	1
Rizzo et al (47)	*150	K	N	N	Y	P	5
Van Poppel et al (58)	312	H	P	P	P	N	5
Wang et al (43)	106	K, B	P	P	P	P	4
Wright et al (52)	60	B	P	P	P	Y	3

^a Initial study sample size refers to the initial size of the sample with respect to individual workers. Where the distinction was permitted, this was the size of the study sample following exclusions on the basis of eligibility, initial inability to contact and initial refusal to participate, but before any loss of sample for reasons of non-response during measurement or withdrawal. Asterisks (*) indicate cases in which numbers were either estimated by the reviewers or were reported as approximate by the authors.

^b Each of the following domains of potential bias was first assessed by several items in the methodological quality assessment form: CSG, II, OA and SA. The reviewers then assessed whether they were confident that the potential for bias in that domain had been minimized: Yes (Y); Partly (P); No (N). Since each outcome was assessed independently, the table shows the strongest of the assessments among all outcomes for each study.

^c Limitations score is derived from the four domain-specific methodological assessments (see Methods) and ranges from 0 (no limitations) to 8 (most limitations).

Table K. Determination of the strength of evidence for each outcome in studies of the relative effectiveness of higher versus lower engagement training [IQR=interquartile range; N=number of studies; NA=not applicable/not available]

Outcome	Summary of body of evidence in table G				Assessment of body of evidence using evidence synthesis algorithm ^a			Strength of evidence ^b
	N		IQR	Median of effect sizes	Number of good or fair studies	Consistency	Median of effect sizes	
	Good	Fair						
Knowledge	0	1	NA	NA	Too few	NA	NA	Insufficient
Attitudes	0	1	NA	0.12	Too few	NA	Sufficient	Insufficient
Behaviors	1	2	0.02–0.08	0.06	Enough	Consistent	Less than sufficient	Insufficient
Health	0	1	NA	0.60	Too few	NA	Large	Insufficient

^a Descriptors indicate the result of assessing a feature of a body of evidence using the evidence synthesis algorithm shown in Table H.

^b The resulting conclusion about strength of evidence following the assessment of a body of evidence against the evidence synthesis algorithm (table H).