RESEARCH

Behavioral skills training for teaching safety skills to mental health service providers compared to training-as-usual: a pragmatic randomized control trial

Elizabeth Lin[1*], Mais Malhas[1], Emmanuel Bratsalis[1], Kendra Thomson[1,2], Fabienne Hargreaves[1], Kayle Donner[1], Heba Baig[1], Rhonda Boateng[1], Rajlaxmi Swain[1], Mary Benisha Benadict[1] and Louis Busch[1]

Abstract

Background Violence in the healthcare workplace has been a global concern for over two decades, with a high prevalence of violence towards healthcare workers reported. Workplace violence has become a healthcare quality indicator and embedded in quality improvement initiatives of many healthcare organizations. The Centre for Addiction and Mental Health (CAMH), Canada's largest mental health hospital, provides all clinical staff with mandated staff safety training for self-protection and team-control skills. These skills are to be used as a last resort when a patient is at imminent risk of harm to self or others. The purpose of this study is to compare the effectiveness of two training methods of this mandated staff safety training for workplace violence in a large psychiatric hospital setting.

Methods Using a pragmatic randomized control trial design, this study compares two approaches to teaching safety skills CAMH's training-as-usual (TAU) using the 3D approach (description, demonstration and doing) and behavioural skills training (BST), from the field of applied behaviour analysis, using instruction, modeling, practice and feedback loop. Staff were assessed on three outcome measures (competency, mastery and confidence), across three time points: before training (baseline), immediately after training (post-training) and one month later (follow-up). This study was registered with the ISRCTN registry on 06/09/2023 (ISRCTN18133140).

Results With a sample size of 99 new staff, results indicate that BST was significantly better than TAU in improving observed performance of self-protection and team-control skills. Both methods were associated with improved skills and confidence. However, there was a decrease in skill performance levels at the one-month follow-up for both methods, with BST remaining higher than TAU scores across all three time points. The impact of training improved staff confidence in both training methods and remained high across all three time points.

Conclusions The study findings suggest that BST is more effective than TAU in improving safety skills among healthcare workers. However, the retention of skills over time remains a concern, and therefore a single training session without on-the-job-feedback or booster sessions based on objective assessments of skill may not be sufficient. Further research is needed to confirm and expand upon these findings in different settings.

*Correspondence:
Elizabeth Lin
Elizabeth.lin@camh.ca
Full list of author information is available at the end of the article

Introduction

Violence in the healthcare workplace has been a global concern for over two decades. In 2002, a joint task force of the International Labour Office (ILO), World Health Organization, Public Services International, and the International Council of Nurses created an initiative to address this issue [1]. One result was the documentation of a high international prevalence of violence towards healthcare workers showing that as many as half or more experienced physical or psychological violence in the previous year [2, 3]. Since then, workplace violence has become a healthcare quality indicator and been embedded in the quality improvement initiatives of many healthcare organizations (for example, Health Quality Ontario [4]). Conceptually, it is also reflected in the expansion of the Triple Aim framework to the Quintuple Aim to include staff work-life experience [5].

Despite these efforts, the high prevalence of workplace violence in healthcare persists [6]. Two meta-analyses, representing 393,344 healthcare workers, found a 19.3% pooled prevalence of workplace violence in the past year among which 24.4% and 42.5% reported physical and psychological violence experiences, respectively [7, 8]. The literature also highlighted that workers in mental health settings were at particular risk [8, 9]. A systematic review of violence in U.S. psychiatric hospitals found between 25 to 85 percent of staff encountering physical aggression in the past year [10]. Partial explanations for this wide range include methodological, population, and setting differences. For example, Gerberich and colleagues [11] surveyed nearly 4,000 Minnesota nurses and found 13 percent reporting physical assault and 38 percent reporting verbal or other non-physical violence in the previous year. Further analyses showed that nurses on psychiatric or behavioral units were twice as likely as those on medical/surgical units to experience physical violence and nearly three times as likely to experience non-physical violence. Ridenour, et al., [12] in a hospital-record study of acute locked psychiatric wards in U.S. Veteran's Hospitals found that 85 percent of nurses had experienced aggression in a 30-day period (85 percent verbal; 81 percent physical). And, in a prospective study of a Canadian psychiatric hospital, Cooper and Mendonca [13] found over 200 physical assaults on nurses within 27 months. While they do not indicate what percentage of nurses were assaulted, their results are consistent with a frequency of between 1 and 2 assaults per week.

Workplace violence has been associated with negative psychological, physical, emotional, financial, and social consequences which impact staff's ability to provide care and function at work [14–16]. A 7-year, population-based, follow-up study in Denmark highlighted the long-term impact of physical and psychological health issues owing to physical workplace violence [17]. Two studies, one in Italy [18] and one in Pakistan [19], have linked workplace violence to demoralization and declining quality of healthcare delivery and job satisfaction among healthcare workers.

Building on these efforts, the ILO published a 2020 report recommending the need for national and organizational work environment policies and workplace training "...on the identified hazards and risks of violence and harassment and the associated prevention and protection measures...." ([20], p. 55). Consequently, many countries [21–23] have committed to creating a safe work environment. In Ontario, Canada, the government has provided guidelines for preventing workplace violence in healthcare [4, 24], and our institution, the Centre for Addiction and Mental Health, launched a major initiative in 2018 to address the physical and psychological safety of patients and staff [25]. A priority component of this initiative is mandatory training for all new clinical staff on trauma-informed crisis prevention, de-escalation skills, and, in particular, safe physical intervention skills [26, 27].

However, the effects of such training, especially for managing aggressive behaviour, are only partially understood. A 2015 systematic review on training for mental health staff [28] and a more recent Cochrane review on training for healthcare staff [29] reported remarkably similar findings. Both noted the inconsistent evidence (due to methodological issues, small numbers of studies, heterogenous results) which made definitive conclusions about the merits and efficacy of training difficult. The more consistent impacts found by Price and colleagues [28] were improved knowledge and staff confidence in their ability to manage aggression. There was some evidence of improved de-escalation skills including the ability to deal with physical aggression [30, 31] and verbal abuse [32]. However, these studies were limited because they used unvalidated scales or simulated, rather than real-world, scenarios. For outcomes such as assault rates, injuries, the incidence of aggressive events, and the use of physical restraints, the findings were mixed or difficult to generalize due to the inconsistent evidence.

Similarly, Geoffrion and colleagues [29] found some positive effect of skills-training on knowledge and attitudes, at least short-term, but noted that support for longer-term effects was less sure. The evidence for impacts on skills or the incidence of aggressive behaviour was even more uncertain. They also noted that the literature was limited because it focused largely on nurses. They concluded, "education combined with training may not have an effect on workplace aggression directed toward healthcare workers, even though education and training may increase personal knowledge and positive attitudes" ([29], p. 2). Among their recommendations were the need to evaluate training in higher-risk settings such as mental healthcare, include other healthcare professionals who also have direct patient contact in addition to nurses, and use more robust study designs. In addition, the literature evaluating training procedures focussed on self-reported rather than objective measures of performance.

Given the concerns with demonstrating effectiveness, the violence prevention literature has tended to focus on training modalities and immediate post-training assessment rather than on skill retention over time. In a systematic review of prevention interventions in the emergency room, Wirth et al. [21] found only five out of 15 included studies that noted any kind of evaluation in the period after training (generally two to nine months post-training) while Geoffrion, et al. [29] identified only two among the nine studies in their meta-analysis that had follow-up skills assessments. However, for both of these reviews, the studies doing follow-up evaluations focused on subjective, self-reported outcomes (empathy, confidence, self-reported knowledge) with no objective behavioral skills measures. Both Wirth et al. [21] and Leach et al. [33] cite studies noting a loss of effectiveness of prevention skills (between three to six months post-training), but specific percentages of retention were not provided.

The present study sought to address these gaps by comparing two approaches to teaching safety skills for managing aggressive patient/client behaviour. The setting was a large psychiatric teaching hospital; the sample was drawn from all new clinical staff attending their mandated on-boarding training; and we used a pragmatic randomized control trial design. In addition, we added a 1-month post-training assessment to evaluate skill retention. Our control intervention was the current training-as-usual (TAU) in which trainers "describe" and "demonstrate", and trainees "do" by practicing the demonstrated skill but without objective checklist-guided performance assessment by the trainer. Our test intervention was behavioural skills training (BST) [34, 35] drawn from the field of applied behaviour analysis [36].

BST is a performance- and competency-based training model that uses an instructional, modeling, practice, and feedback loop to teach targeted skills to a predetermined performance level. Checklists guide the instructional sequence and the determination of whether or not the predetermined performance threshold has been reached. Considerable evidence indicates that BST can yield significant improvement in skills post-training, over time, and across different settings [37–39]. It has been used to train a wide range of participants, including behavior analysts, parents, and educators, to build safety-related skills and manage aggressive behavior [37, 40, 41].

Methods

As previously described [42], our objective was to compare the effectiveness of TAU against BST. Our hypotheses, stated in null form, were that these methods would not differ significantly in:

1. Observer assessment of self-protection and team-control physical skills.
2. Self-assessed confidence in using those skills.

Study participants were recruited from all newly-hired clinical staff attending a mandatory two-week orientation. Staff were required to register beforehand for a half-day, in-person, physical safety skills session. They were randomized to a session at the time of registration, and the sessions were then randomized to TAU or BST. All randomization was performed by RB using GraphPad software [43].

The physical skills training was scheduled for a 3.5 h session on one day of the mandatory onboarding. At the end of the previous day, attendees were introduced to the study (including the fact that it was a randomized study) and asked for consent to email them a copy of the informed consent. On the morning of the physical skills training, a research team member met with attendees to answer questions and then meet privately with each individual to ascertain if they wished to participate and sign the informed consent. The trainers and session attendees were thus unaware of who was or was not in the study. Recruitment began January 2021, after ethics approval, and continued until September 2021 when the target of at least 40 study participants completing all assessments for each training condition was reached. The target sample size was chosen to allow 80-percent power to detect a medium to large effect size [44].

Both methods taught the same 11 target skills for safely responding to patients/clients that may exhibit harm to self or others (e.g., aggressive behaviour) during their hospital admission. These skills, defined by the hospital as mandatory for all newly hired staff, included six

Lin *et al. BMC Health Services Research*   (2024) 24:639

Page 4 of 13

self-protection and five team-control (physical restraint) skills (see Appendix A). Each target skill had defined components and a specific sequence in which they were taught as outlined on performance checklists (see Appendix B for a checklist example).

The two methods differed in how these sequences were administered. For BST, the trainers used the performance checklists to guide the training sequence (instruction, modeling, rehearsal, and feedback) and to indicate when the trainee was ready to move on to the next skill [34] (see Appendix C for BST sequence). In BST, common practice is to define successful performance criteria a priori (e.g., up to three correct, consecutive executions at 100% [45]). However, because the physical skills training session in our study had to be completed within the scheduled 3.5 h, the criterion was lowered for practical reasons to one correct performance (defined as 80% of the components comprising that skill) with the added goal of aiming for up to 5 times in a row if time allowed before moving on to the next skill. In contrast, while TAU included elements of modeling, practice, and feedback, it did not systematically assess skill acquisition nor impose any specific level of success before proceeding to the next skill.

Measures
There were three outcome measures, two observer-based assessments of skill acquisition (competence and mastery) and one self-reported confidence measure. Competence was defined as the percentage of components comprising an individual skill that were correctly executed (e.g., if a skill had 10 components and only six were executed properly, the competence score for that skill would be 60%). Mastery was the threshold defining when a competence score was felt to indicate successful achievement of a skill and to indicate some degree of the durability of the skill acquisition [46]. For our study, we expanded mastery to apply to the two categories of self-protection and team-control (rather than to each individual skill) using the average competence scores for the skills within each category. Mastery was pre-defined as 80-percent, a commonly used threshold [28, 47].

The outcome measures were assessed at three time points: immediately before training (baseline), immediately after training (post-training), and one month later (follow-up). The hospital provided limited descriptive information (professional role, department) for all registrants for administrative purposes but for confidentiality reasons did not provide personal information such as age or gender/sex. The research team elected not to collect personal information for two reasons. First, the primary study concern was to evaluate the main effect of training method rather than developing predictive models,

and the expected result of the randomization process was that potential covariates would not be systematically biased in the two study groups. Second, we would not be able to use this information to compare participants with non-participants to identify biases in who consented to be in the study. We were able to compare them on department role and profession by subtracting the aggregated study-participant information from the aggregated hospital-provided information – the only form of the hospital-provided information available to the research team (see Table 1 below). In addition, since degree of patient contact was an important factor in the likelihood of needing to exercise safety skills, the research team also created an algorithm estimating which combinations of professional role and department were likely to have direct, less direct, or rare/low patient contact.

Participants were also asked at baseline and follow-up how many events they encountered in the previous month that required the use of these skills. This information was collected because of our interest in testing a post-hoc hypothesis that those with actual experience would score higher than those who did not.

All assessments were carried out following a standardized protocol. To ensure that registrants remained blinded to which colleagues were in the study, each registrant's skill acquisition was assessed privately by a research team member at baseline and post-training using the performance checklists. Only assessments for those consenting to participate were videotaped. Study participants were then asked to return one month later for a follow-up assessment which was also videotaped. For the purposes of post-hoc analyses, participants completing all three assessments were defined as 'completers' while those completing baseline and post-training assessments but not the one-month follow-up were 'non-completers.'

The same performance checklists used by the BST trainers were then used by trained observers blinded to the participant's training method to assess the videotapes. As described previously [42], interobserver agreement (IOA) was routinely evaluated throughout the study with the final value being 96% across the 33% of the performance assessment videos scored for the IOA calculation.

Skill acquisition outcomes were calculated using the checklist-based observer assessments of the videotapes. The percentage of correctly executed components for each target skill was established. Then, these percentages were averaged across the six self-protection target skills and across the five team-control target skills to create competence scores. Finally, the predefined threshold of 80% was applied to the competence scores to determine which participants met the mastery threshold [47, 48].

Lin *et al. BMC Health Services Research*   (2024) 24:639

Page 5 of 13

Self-reported confidence was assessed on a 10-point Likert scale ('not at all' to 'extremely' confident) using a version of our institution's standard assessment questions adapted for this study (See Appendix D).

Statistical analysis

R software was used to generate descriptive statistics (frequencies, percentages) and test our hypotheses [49]. Generalized linear mixed models (GLMM) were used to test nested main and interaction effects using likelihood-ratio chi-square statistics for the post-training and follow-up results as there were no baseline differences. GLMM was also used to evaluate BST-TAU differences at the three study time points [50, 51]. For the BST-TAU comparisons, we used Cohen's *d* as a guide for evaluating the practical significance of the differences for the continuous measures (competence, confidence). We used Cohen's suggested thresholds [52] of 0.2, 0.5, and 0.8 for small, medium, and large effect sizes conservatively by applying them to both the point estimates and 95% confidence intervals. Thus, for example, a Cohen's *d* where the confidence interval went below 0.2 would be interpreted as non-meaningful. For the categorical measure of mastery, we used BST-TAU risk ratios. Confidence intervals for all effect size measures were obtained using bootstrapping. Independent-samples *t*-tests were used for the post-hoc analyses and, along with chi-square tests, to compare the completers and non-completers.
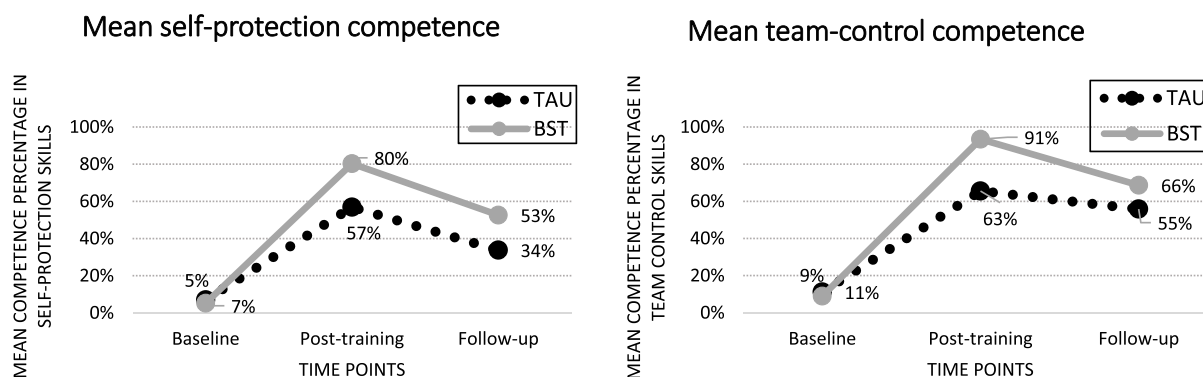
Results

One hundred ninety-nine staff consented to participate in the study out of a total of 360 session attendees (55%). Of these, 108 (54%) had been randomly assigned to a BST session and 91 (46%) to a TAU session. Half ($n = 99$) completed assessments at all three time points (44% TAU; 55% BST). These 99 (hereafter 'study completers') constituted 28 percent of all session attendees.

Among the non-completers, 53 had been assigned to BST and 47 to TAU. Eight were classified as incomplete because of technical software issues when video-recording one of their assessments and one (the first participant) because the IOA process prompted substantive changes to the assessment checklist. The primary reason for the remaining non-completers was missing the follow-up assessment (91 individuals: 50/53 BST, 41/47 TAU) largely due to difficulties scheduling a non-mandatory event during the pandemic (e.g., units restricting staff from leaving because of clinical staff shortages or patient outbreaks, staff illness).

Descriptive information for the expected degree of patient contact and for hospital department is shown in Table 1 for study participants (completers, non-completers), non-participants, and the total group of session attendees. No significant differences were found when comparing participants versus non-participants or study completers versus non-completers in terms of expected patient contact ($\chi^2(2) = 0.36$, n.s.; $\chi^2(2) = 2.22$, n.s.; respectively) or department type ($\chi^2(3) = 4.40$; ($\chi^2(3) = 1.00$, n.s.; respectively).

Table 1 Expected patient contact and department types for study participants, non-participants, and total session attendees

Characteristic	Study Participants ($n = 199$)		Non-Participants ($n = 161$)	Total Session Attendees ($n = 360$)
	Completers $n = 99$	Non-Completers $n = 100$		
Expected Patient Contact: n (%)				
Direct	93 (94)	97 (97)	155 (96)	345 (96)
Nurse	37 (37)	48 (48)	75 (47)	160 (44)
Security	6 (6)	5 (5)	5 (3)	16 (4)
Other	50 (51)	44 (44)	75 (47)	169 (47)
Less Direct	4 (4)	3 (3)	4 (3)	11 (3)
Rare/None	2 (2)	0 (0)	2 (1)	4 (1)
Department Type: n (%)				
Inpatient	57 (58)	64 (64)	82 (51)	203 (56)
Outpatient	9 (9)	9 (9)	23 (14)	41 (11)
Both	30 (30)	26 (26)	53 (33)	109 (30)
Hospital Admin	2 (2)	1 (1)	2 (1)	5 (1)

Lin *et al. BMC Health Services Research*   (2024) 24:639

Page 6 of 13

Mean self-protection competence

Mean team-control competence

Competence: Linear Mixed Model (study completers, n=99)							
	Self-protection				**Team-control**		
	Df	χ^2	**p-value**		**Df**	χ^2	**p-value**
Method	1	34.46	<.001		1	50.42	<.001
Time	1	81.29	<.001		1	56.51	<.001
Method-by-Time	1	0.97	n.s.		1	17.41	<.001
TIME POINT	**BST-TAU**	**Cohen's d [95% CI]**	**Effect size**		**BST-TAU**	**Cohen's d [95% CI]**	**Effect size**
Post-training	0.23	1.45 [1.02,1.87]	large		0.28	2.55 [2.08, 3.02]	large
Follow-up	0.19	0..82 [0.40, 1.23]	small		0.12	0.62 [0.21, 1.03]	small

Fig. 1 Observer-rated self-protection and team-control competence skills in TAU and BST across time-points

Figure 1 depicts the self-protection and team-control competence scores for the study completers (left and right sides, respectively). The hypothesis-testing results showed a significant difference by training Method (self-protection: $\chi^2(1) = 34.46$, $p < 0.001$; team-control: $\chi^2(1) = 50.42$, $p < 0.001$). There was also a significant decline between post-training and follow-up (Time) for both skill categories independent of Method (self-protection: $\chi^2(1) = 81.29$, $p < 0.001$; team-control: $\chi^2(1) = 56.51$, $p < 0.001$), and a significant Method-by-Time interaction independent of Method and Time for team-control skills ($\chi^2(1) = 17.41$, $p < 0.001$). BST-TAU comparisons showed no difference at baseline for either type of skill (not shown). However, BST was significantly better than TAU at both post-training (self-protection: Cohen's $d = 1.45$ [1.02, 1.87], large effect size; team-control: Cohen's $d = 2.55$ [2.08, 3.02]; large effect size) and follow-up (respectively – Cohen's $d = 0.82$ [0.40, 1.23]; Cohen's $d = 0.62$ [0.21, 1.03], both small effect sizes). For both methods, competence scores dropped between post-training and follow-up although not to the original baseline levels.

The skill mastery results for the study completers are shown in Fig. 2. The mastery patterns paralleled the competence patterns in that BST was significantly better than TAU (self protection: $\chi^2(1) = 28.82$, $p < 0.001$; team-control: $\chi^2(1) = 72.87$, $p < 0.001$). There was also a significant Time effect independent of Method (self-protection: $\chi^2(1) = 27.54$, $p < 0.001$; team-control: $\chi^2(1) = 33.03$, $p < 0.001$). There were no significant interactions for either type of skill once the effects of Method and Time were accounted for. BST-TAU comparisons showed no difference in percent achieving Mastery at baseline (not shown) but large risk ratios at both post-training (self-protection: 13.43 [4.01, >1000]; team-control: 31.24 [8.45, >1000] and follow-up [self-protection: 12.30 [1.58, >1000]; team-control: 30.60 [6.75. >1000]).

Confidence scores for the study completers are shown in Fig. 3. The only significant main effect was for Time (self-protection: $\chi^2(1) = 36.87$, $p < 0.001$; team-control: $\chi^2(1) = 21.08$, $p < 0.001$). For both skill categories, the scores increased between baseline and post-training and then dropped at follow-up but not to the original baseline levels.

Lin *et al. BMC Health Services Research*   (2024) 24:639

Page 7 of 13

% self-protection mastery

% team-control mastery

Mastery: Generalized Linear Mixed Model (study completers, n=99)

	Self-protection			Team Control		
	Df	χ^2	p-value	Df	χ^2	p-value
Method	1	28.82	<.001	1	72.87	<.001
Time	1	27.54	<.001	1	33.03	<.001
Method-by-Time	1	0.50	n.s.	1	1.89	n.s.

TIME POINT	BST/TAU risk ratio[1]	95% CI[2]	BST/TAU risk ratio[Error! Bookmark not defined.]	95% CI[Error! Bookmark not defined.]
Post-training	13.43	4.01, >1000	**31.24**	8.45, >1000
Follow-up	12.30	1.58, >1000	**30.60**	6.75, >1000

[1] Risk ratios based on model estimates rather than observed proportions

[2] Wide CIs because of cell proportions near zero in TAU condition

Fig. 2 Observer-rated self-protection and team-control mastery (Predefined as 80% or better competence) by TAU and BST across time-points

To assess what impact the high no-show rate for the one-month follow-up could have had, we compared the completers and the non-completers on the six post-training outcomes (competence, mastery, and confidence for self-protection and for team-control). Non-completers had slightly lower scores than completers except for the two confidence measures where their self-assessments were higher (not shown). However, the only significant difference between the two groups was for self-protection competence means (0.70 vs 0.63, completers vs non- completers, $t(195) = 2.40$, $p = 0.017$).

In terms of past-month experience, few study completers reported events requiring self-protection (19 at baseline, 9 at follow-up) or team-control skills (14 at baseline, 14 at follow-up). Consequently, we only examined the presence or absence of experience without breaking it down by training method. We found non-significant results for both competence and mastery (not shown) but a potential impact on confidence for self-protection skills at follow-up and for team-control skills at baseline and post-training (Fig. 4).

4. Summary and discussion.

Our strongest finding was that BST was significantly better than TAU in improving the observed performance of self-protection and team-control skills. While follow-up scores decreased for both methods, BST scores remained higher than TAU scores. The impact of training on staff confidence differs from these patterns in that confidence scores improved noticeably at post-training and remained relatively high at follow-up. Further, our post-hoc analyses suggested that recent experience using safety skills might have a greater impact on confidence than on observed skill performance. We also found that training, regardless of method, was independently associated with improved observer-scored skills and self-reported confidence.

Mean self-protection confidence

Mean team-control confidence
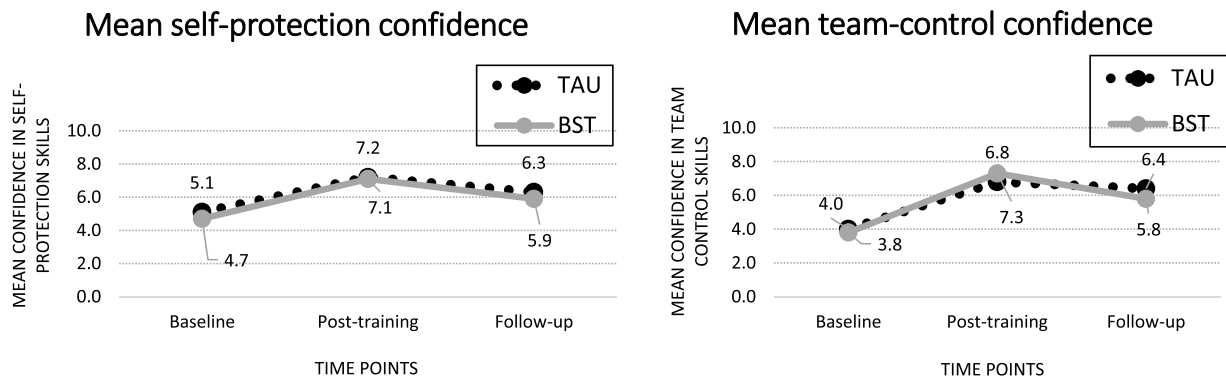
Confidence: Linear Mixed Model (study completers, n=99)							
	Self-protection				Team Control		
	Df	χ^2	p-value		Df	χ^2	p-value
Method	1	0.56	n.s.		1	0.08	n.s.
Time	1	36.87	<.001		1	21.08	<.001
Method-by-Time	1	1.20	n.s.		1	6.81	0.009
TIME POINT	**BST-TAU**	**Cohen's d [95% CI]**	**Effect size**		**BST-TAU**	**Cohen's d [95% CI]**	**Effect size**
Post-training	-0.11	0.06 [-0.19,0.31]	n.s.		0.41	0.19 [-0.13, 0.51]	n.s.
Follow-up	-0.45	0.22 [-0.11, 0.56]	n.s.		-0.68	0.30 [-0.07, 0.66]	n.s.

Fig. 3 Self-rated self-protection and team-control confidence in TAU and BST across time-points
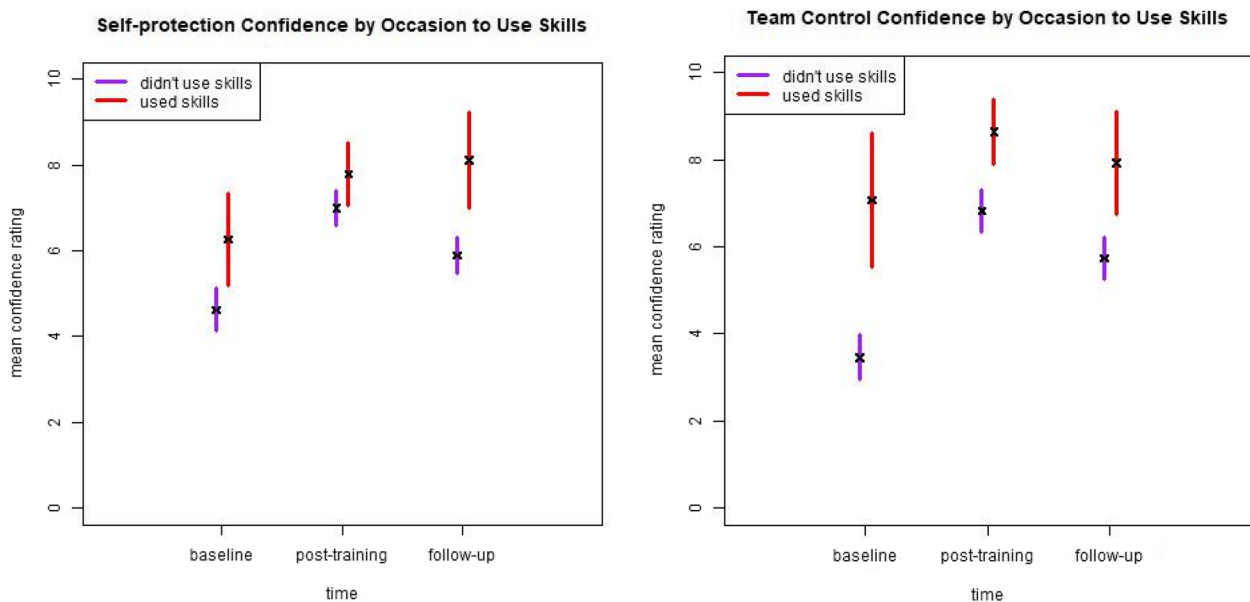
Fig. 4 Self-rated self-protection and team-control confidence by occasion to use skills in the past month across time-points

Lin *et al. BMC Health Services Research*   (2024) 24:639

Page 9 of 13

The better performance of BST is consistent with the fact that it incorporates training elements that are supported both by current educational and learning theories and evidence of effectiveness [46, 53–55]. While both BST and TAU can be considered 'outcomes based' [54], the key difference is the BST's use of the checklist. Based directly on the desired behavioral outcomes, this tool simultaneously creates a common understanding because it is shared with the trainees, ensures consistent and systematic training across all BST trainees, pinpoints where immediate and personalized feedback is needed to either correct or reinforce performance, and tracks the number of correct repetitions required to meet mastery criteria as well as support retention [46, 56, 57]. By contrast, TAU does not use a checklist and the kind and amount of feedback or practice repetitions is left to the trainer's discretion.

However, there are at least two questions regarding whether BST produced the expected results. The BST framework requires continued rehearsal and feedback until a specified performance criterion is reached [34]. However, our mandatory safety training had practical, unmodifiable constraints. The institution required the safety-training sessions be completed in 3.5 h which meant that BST trainers were limited in their ability to use the more stringent performance criteria described in the literature. For example, it was not practical to set the performance criterion at higher than 80 percent. In addition, all BST completers were able to demonstrate 80-percent correct performance for each skill at least once, but not all were able to demonstrate five consecutive, correct executions within the allotted time. If the requirement of five in a row at 80% or higher had been implemented, then the post-training scores (and potentially the 1-month follow-up scores) for the BST completers could have been higher.

A second question is what level of skill retention should be expected at follow-up. The BST scores at one-month follow-up constituted 66% and 73% of the competence scores at post-training (self-protection and team-control, respectively) and 30% and 41% of the mastery percentages at post-training (self-protection and team-control, respectively). Although BST and elements of performance feedback models have been found to be effective in staff training with successful retention over time [58–62], finding appropriate comparators for our study was challenging because there are no studies where BST has been used for training such a large and diverse group of staff. Further, as noted above, the body of workplace violence prevention literature has not consistently focussed on retention. However, the broader training and education literature does suggest that our results are consistent with or somewhat lower than those from other studies.

Offiah et al. [63] found that 45 percent of medical students retained the full set of clinical skills 18 months after completing simulation training, and Bruno and colleagues [64] found published retention rates ranging between 75 and 85 percent across time periods between four to 24 months and across diverse disciplinary fields. Regardless of the comparators, the loss in skill performance after one-month post-training is a concern.

Our interpretation is that reliance on a single session, even with highly structured and competency-based methods, is not adequate particularly in the context of managing distressing events. Efforts should be made to allow for flexibility with respect to setting higher thresholds for success despite organizational restraints for staff training. Furthermore, settings that require these skills to be performed more reliably for both patient and staff safety (e.g., emergency departments, acute care settings, security services) should consider on-the-job feedback or booster sessions based on objective assessments of skill rather than on pre-set amounts of time (e.g. annual refresher). This would be more consistent with the BST literature, as on-the-job training should occur based on an evidence-based approach.

Our finding of a differential impact of training on confidence versus demonstrable skills is consistent with a long-standing, substantial body of research examining the relationship between self-assessment and objective measures of learning [28, 65, 66]. The pattern of non-existent, weak, or even inverse relationships between the two has been shown for a variety of medical staff trainee and education learner groups [28, 29, 67–72]. Consequently, many researchers recommend either not using self-assessments at all or at least ensuring that objective measures are also collected (e.g.,[64, 65]).

The literature does offer some hypotheses for why this discrepancy occurs and, further, why self-assessment continues to be used in medical education and training despite the robust evidence that it does not accurately reflect learning. Katowa-Mukwato and Banda [70] in a study of Zambian medical students suggest that fear of revealing their weaknesses led to a negative correlation between self- and objective-ratings. Persky, et al. [69] reference the theory of 'metacognition' – defined as 'thinking about thinking' (p. 993, [69] – and the 'Dunning-Kruger' effect that the ability to recognize competence (i.e., accurate metacognition) is unevenly distributed. There is also discussion as to why these measures continue to be used and suggestions of how best to use them. Yates et al. [65] suggest that ease of collecting this information is a factor. More complex and nuanced explanations are offered by Lu, et al. [66] and Tavares, et al. [73] who note that self-assessment is an important component in theories of learning and evaluation and

that self-perception and self-reflection (particularly when objective findings are shared) are critical ingredients for supporting medical and continuing profession education in a self-regulating profession.

Because the goal of our study was to assess the effectiveness of two training methods, we did not collect information or have the opportunity to explore any of these potential reasons for why self-reported and objective measures are discrepant or to evaluate the best use of that discrepancy. The modest contributions that our study adds are that selecting a higher-risk setting, including non-nursing healthcare professionals, using a more rigorous study design (as recommended by Geoffrion, et al. [29]), and attempting to account for recent experience do not appear to alter this pattern.

The major strength of our study is its design. Currently, we have identified only one other study evaluating the impact of BST training for clinical staff using a randomized control trial design [41]. Other strengths are our inclusion of a large percentage of non-nursing, direct-care staff, our use of both self-reported and observer-assessed outcome measures, and our findings regarding retention. These strengths allow us to add to the evidence base already established in the literature.

However, interpretation of our results should consider several limitations. Conducting a research study on full-time clinical staff during a pandemic meant that a high percentage of those consenting to be in the study did not complete their 1-month follow-up assessment. The reported reasons for missing the third assessment (unit restrictions or short staffing because of the pandemic) are consistent with the demographic differences between completers and non-completers in that they were more likely to be nurses or working on inpatient units. Our comparison of the post-training scores of the completers and non-completers suggested that the no-shows had slightly lower post-training observed skill performance (but slightly better confidence ratings). If we had managed to assess the non-completers at follow-up, our reported findings may have been diluted although it is unlikely that this would have completely negated the large effect sizes.

The time constraints on the mandatory training meant that we were unable to fully apply either the BST mastery criteria commonly reported in the literature (i.e., three correct, consecutive executions [28, 47] or the one we would have preferred (i.e., five correct executions). While this type of limitation is consistent with the pragmatic nature of our design, it likely had an impact on our findings in terms of potentially lowering the post-training BST competency and mastery scores and, perhaps more importantly, contributing to the lower retention rates at 1-month follow-up [56].

The 45-percent refusal rate by the training registrants is another concerning issue. Anecdotal reports from the training team were that the response rate was very low at the start of the study because many of the new hires were nervous about being videotaped (a specific comment reported was that it reminded some of the new graduates of 'nursing school.') and were unsure of the purpose of the study. The team then changed to a more informal, conversational introduction describing the need for the study as well as reassuring attendees that it was the training, not the participants, that was being evaluated. The team's impression was that this improved the participation rate. The participants and non-participants were not statistically different in terms of their expected patient contact and department role. However, we cannot preclude that there may have been systematic biases for other unmeasured characteristics.

Another limitation, as identified by Price, et al. [28], is that we used artificial training scenarios, though this may be unavoidable given the low frequency of aggressive events and the ethics of deliberately exposing staff to these events. Also, we only measured the skills directly related to handling client/patient events. We were not able to access information on event frequency or severity, staff distress and complaints, or institutional-level measures such as lost workdays due to sick leave, staff turnover, or expenditures [29, 33]. A further gap, which is important but difficult to assess, is whether there is any impact of staff safety training on the clients or patients who are involved.

Given these strengths and limitations, we see our study as adding one piece of evidence that needs to be a) confirmed or disconfirmed by other researchers in both the same and different settings and b) understood as part of a complex mix of ingredients. Specific areas for further research arising directly out of our findings include evaluating whether less constrained training time would improve attainment of skill mastery, exploration and evaluation of methods to increase skill retention over time, and, most importantly but also more difficult to assess, the impact on patients and clients of staff safety skills training. More evidence on these fronts will hopefully contribute to maintaining and improving workplace safety.

Supplementary Information

> **Additional file 1.**
>
> **Additional file 2.**

Lin *et al. BMC Health Services Research*   (2024) 24:639

Page 11 of 13

Availability of data and materials
The dataset generated and analysed during the current study is not publicly available due to the fact that it is part of a larger internal administrative data collection but is available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
This study was approved by the Research Ethics Board of the Centre for Addiction and Mental Health (#101/2020). Informed consent was obtained from all subjects participating in the study. All interventions were performed in accordance with the Declaration of Helsinki. This study was registered with the ISRCTN registry on 06/01/2023 (ISRCTN18133140).

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Education, Centre for Addiction and Mental Health, Toronto, ON, Canada. [2]Department of Applied Disability Studies, Brock University, St. Catharines, ON, Canada.

References
1. International Labour Organization. Joint Programme Launches New Initiative Against Workplace Violence in the Health Sector. 2002. Available from: https://www.ilo.org/global/about-the-ilo/newsroom/news/WCMS_007817/lang--en/index.htm. Cited 2022 Aug 5.
2. Di Martino V. Workplace violence in the health sector. Country case studies Brazil, Bulgaria, Lebanon, Portugal, South Africa, Thailand and an Additional Australian Study. Geneva: ILO/ICN/WHO/PSI Joint Programme on Workplace Violence in the Health Sector; 2002.
3. Needham I, Kingma M, O'Brien-Pallas L, McKenna K, Tucker R, Oud N, editors. Workplace Violence in the Health Sector. In: Proceedings of the First International Conference on Workplace Violence in the Health Sector - Together, Creating a Safe Work Environment. The Netherlands: Kavanah; 2008. p. 383.
4. Health Quality Ontario. Quality Improvement Plan Guidance: Workplace Violence Prevention. 2019.
5. Bodenheimer T, Sinsky C. From triple to quadruple aim: care of the patient requires care of the provider. Ann Fam Med. 2014;12(6):573–6.
6. Somani R, Muntaner C, Hillan E, Velonis AJ, Smith P. A Systematic review: effectiveness of interventions to de-escalate workplace violence against nurses in healthcare settings. Saf Health Work. 2021;12(3):289–95. https://doi.org/10.1016/j.shaw.2021.04.004.
7. Li Y, Li RQ, Qiu D, Xiao SY. Prevalence of workplace physical violence against health care professionals by patients and visitors: a systematic review and meta-analysis. Int J Environ Res Public Health. 2020;17(1):299.
8. Liu J, Gan Y, Jiang H, Li L, Dwyer R, Lu K, et al. Prevalence of workplace violence against healthcare workers: a systematic review and meta-analysis. Occup Environ Med. 2019;76(12):927–37.
9. O'Rourke M, Wrigley C, Hammond S. Violence within mental health services: how to enhance risk management. Risk Manag Healthc Policy. 2018;11:159–67.
10. Odes R, Chapman S, Harrison R, Ackerman S, Hong OS. Frequency of violence towards healthcare workers in the United States' inpatient psychiatric hospitals: a systematic review of literature. Int J Ment Health Nurs. 2021;30(1):27–46.
11. Gerberich SG, Church TR, McGovern PM, Hansen HE, Nachreiner NM, Geisser MS, et al. An epidemiological study of the magnitude and consequences of work related violence: the minnesota nurses' study. Occup Environ Med. 2004;61(6):495–503.
12. Ridenour M, Lanza M, Hendricks S, Hartley D, Rierdan J, Zeiss R, et al. Incidence and risk factors of workplace violence on psychiatric staff. Work. 2015;51(1):19–28.
13. Cooper AJ, Mendonca JD. A prospective study of patient assaults on nurses in a provincial psychiatric hospital in Canada. Acta Psychiatr Scand. 1991;84(2):163–6.
14. Hiebert BJ, Care WD, Udod SA, Waddell CM. Psychiatric nurses' lived experiences of workplace violence in acute care psychiatric Units in Western Canada. Issues Ment Health Nurs. 2022;43(2):146–53.
15. Lim M, Jeffree M, Saupin S, Giloi N, Lukman K. Workplace violence in healthcare settings: the risk factors, implications and collaborative preventive measures. Ann Med Surg. 2022;78:103727.
16. Lanctôt N, Guay S. The aftermath of workplace violence among healthcare workers: a systematic literature review of the consequences. Aggress Violent Behav. 2014;19(5):492–501.
17. Friis K, Pihl-Thingvad J, Larsen FB, Christiansen J, Lasgaard M. Long-term adverse health outcomes of physical workplace violence: a 7-year population-based follow-up study. Eur J Work Organ Psychol. 2019;28(1):101–9.
18. Berlanda S, Pedrazza M, Fraizzoli M, de Cordova F. Addressing risks of violence against healthcare staff in emergency departments: the effects of job satisfaction and attachment style. Biomed Res Int. 2019;28(5430870):1–12.
19. Baig LA, Ali SK, Shaikh S, Polkowski MM. Multiple dimensions of violence against healthcare providers in Karachi: results from a multicenter study from Karachi. J Pakistani Med Assoc. 2018;68(8):1157–65.
20. International Labour Organization. Safe and healthy working environments free from violence and harassment. International Labour Organization. Geneva: International Labour Organization; 2020. Available from: https://www.ilo.org/global/topics/safety-and-health-at-work/resources-library/publications/WCMS_751832/lang--en/index.htm.
21. Wirth T, Peters C, Nienhaus A, Schablon A. Interventions for workplace violence prevention in emergency departments: a systematic review. Int J Environ Res Public Health. 2021;18(16):8459.
22. U.S. Congress. Workplace Violence Prevention for Health Care and Social Service Workers Act. U.S.A.; 2020. Available from: https://www.congress.gov/bill/116th-congress/house-bill/1309.

23. Lu A, Ren S, Xu Y, Lai J, Hu J, Lu J, et al. China legislates against violence to medical workers. TneLancet Psychiatry. 2020;7(3):E9.
24. Government of Ontario Ministry of Labour T and SD. Preventing workplace violence in the health care sector | ontario.ca. 2020. Available from: https://www.ontario.ca/page/preventing-workplace-violence-health-care-sector. Cited 2020 Jul 28.
25. Safe and Well Committee. Safe & Well Newsletter. 2018. Available from: http://insite.camh.net/files/SafeWell_Newsletter_October2018_107833.pdf.
26. Patel MX, Sethi FN, Barnes TR, Dix R, Dratcu L, Fox B, et al. Joint BAP NAPICU evidence-based consensus guidelines for the clinical management of acute disturbance: de-escalation and rapid tranquillisation. J Psychiatr Intensive Care. 2018;14(2):89–132.
27. Heckemann B, Zeller A, Hahn S, Dassen T, Schols JMGA, Halfens RJG. The effect of aggression management training programmes for nursing staff and students working in an acute hospital setting. A narrative review of current literature. Nurse Educ Today. 2015;35(1):212–9.
28. Price O, Baker J, Bee P, Lovell K. Learning and performance outcomes of mental health staff training in de-escalation techniques for the management of violence and aggression. Br J Psychiatry. 2015;206:447–55.
29. Geoffrion S, Hills DJ, Ross HM, Pich J, Hill AT, Dalsbø TK, et al. Education and training for preventing and minimizing workplace aggression directed toward healthcare workers. Cochrane Database Syst Rev. 2020;9(Art. No.: CD011860). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8094156/. Cited 2021 Dec 3.
30. Paterson B, Turnbull J, Aitken I. An evaluation of a training course in the short-term management of violence. Nurse Educ Today. 1992;12(5):368–75.
31. Rice ME, Helzel MF, Varney GW, Quinsey VL. Crisis prevention and intervention training for psychiatric hospital staff. Am J Community Psychol. 1985;13(3):289–304.
32. Wondrak R, Dolan B. Dealing with verbal abuse: evaluation of the efficacy of a workshop for student nurses. Nurse Educ Today. 1992;12(2):108–15.
33. Leach B, Gloinson ER, Sutherland A, Whitmore M. Reviewing the evidence base for de-escalation training: a rapid evidence assessment. RAND research reports. RAND Corporation; 2019. Available from: https://www.rand.org/pubs/research_reports/RR3148.html. Cited 2021 Nov 26.
34. Parsons MB, Rollyson JH, Reid DH. Evidence-based staff training: a guide for practitioners. Behav Anal Pract. 2012;5(2):2–11.
35. Miltenberger RG, Flessner C, Gatheridge B, Johnson B, Satterlund M, Egemo K. Evaluation of behavioral skills training to prevent gun play in children. J Appl Bahavior Anal. 2004;37:513–6.
36. Baer DM, Wolf MM, Risley TR. Some still-current dimensions of applied behavior analysis. J Appl Behav Anal. 1987;20(4):313–27.
37. Dillenburger K. Staff training. Handbook of treatments for autism spectrum disorder. In: Matson JL, editor. Handbook of treatments for autism spectrum disorder. Switzerland: Springer Nature; 2017. p. 95–107.
38. Kirkpatrick M, Akers J, Rivera G. Use of behavioral skills training with teachers: a systematic review. J Behav Educ. 2019;28(3):344–61.
39. Sun X. Behavior skills training for family caregivers of people with intellectual or developmental disabilities: a systematic review of literature. Int J Dev Disabil. 2020;68(3):247-73.
40. Davis S, Thomson K, Magnacca C. Evaluation of a caregiver training program to teach help-seeking behavior to children with autism spectrum disorder. Int J Dev Disailities. 2020;66(5):348–57.
41. Gormley L, Healy O, O'Sullivan B, O'Regan D, Grey I, Bracken M. The impact of behavioural skills training on the knowledge, skills and well-being of front line staff in the intellectual disability sector: a clustered randomised control trial. J Intellect Disabil Res. 2019;63(11):1291–304.
42. Lin E, Malhas M, Bratsalis E, Thomson K, Boateng R, Hargreaves F, et al. Behavioural skills training for teaching safety skills to mental health clinicians: A protocol for a pragmatic randomized control trial. JMIR Res Protoc. 2022;11(12):e39672.
43. GraphPad. Randomly assign subjects to treatment groups. Available from: https://www.graphpad.com/quickcalcs/randomize1.cfm. Cited 2023 June 6.
44. Van Voorhis CRW, Morgan BL. Understanding power and rules of thumb for determining sample sizes. Tutor Quant Methods Psychol. 2007;3(2):43–50.

45. Erath TG, DiGennaro Reed FD, Sundermeyer HW, Brand D, Novak MD, Harbison MJ, et al. Enhancing the training integrity of human service staff using pyramidal behavioral skills training. J Appl Behav Anal. 2020;53(1):449–64.
46. Wong KK, Fienup DM, Richling SM, Keen A, Mackay K. Systematic review of acquisition mastery criteria and statistical analysis of associations with response maintenance and generalization. Behav Interv. 2022;37(4):993–1012.
47. Richling SM, Williams WL, Carr JE. The effects of different mastery criteria on the skill maintenance of children with developmental disabilities. J Appl Behav Anal. 2019;52(3):701–17 (Available from: https://pubmed.ncbi.nlm.nih.gov/31155708/). Cited 2022 Oct 14.
48. Pitts L, Hoerger ML. Mastery criteria and the maintenance of skills in children with developmental disabilities. Behav Interv. 2021;36(2):522–31. Available from: https://doi.org/10.1002/bin.1778. Cited 2022 Oct 17.
49. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2021.
50. Pinheiro J, Bates D. Mixed-effects models in S and S-PLUS. New York: Springer Science+Business Media; 2006.
51. Bosker R, Snijders TA. Multilevel analysis: An introduction to basic and advanced multilevel modeling. In: Analysis M, editor. London. UK: Sage Publishers; 2012. p. 1–368.
52. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
53. Taylor DCM, Hamdy H. Adult learning theories: Implications for learningand teaching in medical education: AMEEGuide No. 83. Med Teach. 2013;35(11):e1561-72.
54. Nodine TR. How did we get here? A brief history of competency-based higher education in the United States. Competency Based Educ. 2016;1:5–11.
55. Novak MD, Reed FDD, Erath TG, Blackman AL, Ruby SA, Pellegrino AJ. Evidence-Based Performance Management: Applying behavioral science to support practitioners. Perspect Behav Sci. 2019;42:955–72.
56. Fienup DM, Broadsky J. Effects of mastery criterion on the emergence of derived equivalence relations. J Appl Behav Anal. 2017;40:843–8.
57. Fuller JL, Fienup DM. A preliminary analysis of mastery criterion level: effects on response maintenance. Behav Anal Pract. 2018;11(1):1–8.
58. Alavosius MP, Sulzer-Azaroff B. The effects of performance feedback on the safety of client lifting and transfer. J Appl Behav Anal. 1986;19:261–7.
59. Hogan A, Knez N, Kahng S. Evaluating the use of behavioral skills training to improve school staffs' implementation of behavior intervention plans. J Behav Educ. 2015;24(2):242–54.
60. Nabeyama B, Sturmey P. Using behavioral skills training to promote safe and correct staff guarding and ambulation distance of students with multiple physical disabilities. J Appl Behav Anal. 2010;43(2):341–5.
61. Parsons MB, Rollyson JH, Reid DH. Teaching practitioners to conduct behavioral skills training: a pyramidal approach for training multiple human service staff. Behav Anal Pract. 2013;6(2):4–16.
62. Sarakoff RA, Sturmey P. The effects of behavioral skills training on staff implementation of discrete-trial teaching. J Appl Behav Anal. 2004;37(4):535–8.
63. Offiah G, Ekpotu LP, Murphy S, Kane D, Gordon A, O'Sullivan M, et al. Evaluation of medical student retention of clinical skills following simulation training. BMC Med Educ. 2019;19:263.
64. Bruno P, Ongaro A, Fraser I. Long-term retention of material taught and examined in chiropractic curricula: its relevance to education and clinical practice. J Can Chiropr Assoc. 2007;51(1):14–8.
65. Yates N, Gough S, Brazil V. Self-assessment: With all its limitations, why are we still measuring and teaching it? Lessons from a scoping review. Med Teach. 2022;44(11):1296–302. https://doi.org/10.1080/0142159X.2022.2093704.
66. Lu FI, Takahashi SG, Kerr C. Myth or reality: self-assessment is central to effective curriculum in anatomical pathology graduate medical education. Acad Pathol. 2021;8:23742895211013530.
67. Magnacca C, Thomson K, Marcinkiewicz A, Davis S, Steel L, Lunsky Y, et al. A telecommunication model to teach facilitators to deliver acceptance and commitment training. Behav Anal Pract. 2022;15(3):752–752.
68. Naughton CA, Friesner DL. Comparison of pharmacy students' perceived and actual knowledge using the pharmacy curricular outcomes assessment. Am J Pharm Educ. 2012;76(4):63.
69. Persky AM, Ee E, Schlesselman LS. Perception of learning versus performance as outcome measures of educational research. Am J Pharm. 2020;84(7):993–1000.

Lin *et al. BMC Health Services Research*   (2024) 24:639

Page 13 of 13

70. Katowa-Mukwato P, Sekelani SB. Self-perceived versus objectively measured competence in performing clinical pratical procedures by final medical students. Int J Med Educ. 2016;7:122–9.

71. Barsuk JH, Cohen ER, Feinglass J, McGaghie WC, Wayne DB. Residents' procedural experience does not ensure competence: a research synthesis. J Grad Med Educ. 2017;9(2):201–8.

72. Choudhry NK, Fletcher RH, Soumerai SB. Systematic review: the relationship between clinical experience and quality of health care. Ann Intern Med. 2005;142(4):260–73.

73. Tavares W, Sockalingam S, Valanci S, Giuliani M, Davis D, Campbell C, et al. Performance Data Advocacy for Continuing Professional Development in Health Professions. Acad Med. 2024;99(2):153-8. https://doi.org/10.1097/ACM.0000000000005490.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.