European Journal of Behavior Analysis

ISSN: 1502-1149 (Print) 2377-729X (Online) Journal homepage: https://www.tandfonline.com/loi/rejo20

Comparing a student active learning format to equivalence-based instruction

Hanne Augland, Torunn Lian & Erik Arntzen

To cite this article: Hanne Augland, Torunn Lian & Erik Arntzen (2020): Comparing a student active learning format to equivalence-based instruction, European Journal of Behavior Analysis, DOI: 10.1080/15021149.2020.1752513

To link to this article: https://doi.org/10.1080/15021149.2020.1752513

Published online: 13 May 2020.

EUROPEAN JOURNAL OF BEHAVIOR ANALYSIS
https://doi.org/10.1080/15021149.2020.1752513

ARTICLE

Comparing a student active learning format to equivalence-based instruction

Hanne Augland, Torunn Lian and Erik Arntzen

Department of Behavioral Science, Oslo Metropolitan University, Oslo, Norway

ABSTRACT
The present study compared the effectiveness of a student active learning format (SALF) to equivalence-based instruction (EBI), in teaching behavior analytic terms. The EBI condition included matching-to-sample, and SALF included elements from interteaching. Participants experienced both SALF and EBI conditions. Two classes consisting of 48 and 33 participants were assigned to two groups. One group experienced EBI condition in an early phase of the course, while the other group, at the end of the course. The EBI and SALF conditions show to be equally effective. However, EBI was completed in less time than the SALF condition. Participants who met the criterion for stimulus equivalence had a higher score on two different tests for generalization. The results replicate earlier findings in that EBI proves to be effective in teaching concepts in college students. Furthermore, the present results extend previous findings by proving two student active learning formats to be equally effective.

ARTICLE HISTORY
Received 1 April 2019
Accepted 24 March 2020

KEYWORDS
Stimulus equivalence; equivalence-based instruction; student active learning format; college students

Stimulus equivalence is described as stimulus classes where the members in the class are mutually interchangeable stimuli (Green & Saunders, 1998) and refers to stimulus classes in which class membership is not based on physical similarities between the members of the class but the relations between the stimuli. Members in equivalence classes share the properties of reflexivity, symmetry, and transitivity (Sidman & Tailby, 1982). In research on stimulus equivalence, these properties are usually probed for after the establishment of at least two conditional discriminations, sharing one stimulus. That is, the defining features of equivalence classes should emerge without direct training. For example, if selecting B1 in the presence of A1 and selecting C1 in the presence of B1 is established through direct training, then selecting A1 in the presence of B1 without a direct reinforcement history would be an example of symmetry. Selecting C1 in the presence of A1 would be an example of transitivity, and selecting A1 in the presence of C1 would be an example of an equivalence property. Finally, reflexivity requires that each stimulus is selected in the presence of itself. That is selecting A1 comparison stimulus in the presence of A1 sample stimulus.

Research on stimulus equivalence originated in an applied setting in which Sidman and colleagues (Sidman, 1971; Sidman & Cresson, 1973) aimed at teaching patients

CONTACT Hanne Augland augland.hanne@gmail.com

reading comprehension. Since then, a large part of the research on stimulus equivalence has been done within basic research. Nevertheless, procedures based on stimulus equivalence technology have demonstrated the effects of conditional discrimination training and testing for equivalence class formation in applied settings. In recent years, the term equivalence-based instruction (EBI) has been used as a collective term for stimulus equivalence technology and this term will be used throughout the manuscript.

EBI has shown to be effective in generating concept formation in a variety of topics in a wide range of participants. For example, EBI have shown to effectively establish real-world concepts in pre-school children (Barron et al., 2018) reading and reading comprehension (De Souza et al., 2009) and fraction – decimal relations (Leader & Barnes-Holmes, 2001; Lynch & Cuvo, 1995) in children in primary and secondary education. Other researchers have used equivalence procedures to teach addition and subtraction (Henklain & Carmo, 2013) braille reading to children with degenerative visual impairments (Toussaint & Tiger, 2010), coin values to participants diagnosed as mentally retarded (McDonagh et al., 1984), geography skills to adolescents with Fragile X-syndrome (Hall et al., 2006) and autism (LeBlanc et al., 2003), geometry skills in children with high functioning autism (Dixon et al., 2016) and piano skills in a boy with autism (Arntzen et al., 2010) and in college students (Griffith et al., 2018).

Further, researchers have used EBI to teach concepts in higher education. For example, brain anatomy and function (Fienup et al., 2010, 2016, 2015; Pytte & Fienup, 2012), statistical variability (Albright et al., 2015), drawing conclusions in hypothesis testing (Fienup & Critchfield, 2011), identification of logical fallacies (Ong et al., 2018), trigonometry skills (Ninness et al., 2006, 2009), and interpretations of operant functions of behavior (Albright et al., 2016). Taken together, these studies indicate that EBI can be used to teach a variety of concepts in higher education. One of the most persuasive arguments for the effectiveness of EBI is the fact that equivalence classes emerge for most of the participants in these studies.

To our knowledge, only a few studies have investigated the relative effects of EBI and other teaching formats (Fields et al., 2009; Fienup & Critchfield, 2011; Lovett et al., 2011; Ong et al., 2018). A study by Fields et al. (2009) showed that all participants who received EBI formed equivalence classes and that relations generalized to novel exemplars in a paper and pencil posttest. Students receiving EBI showed a higher mean score on the paper and pencil posttest than did participants in the control group. Fienup and Critchfield (2011) found that EBI was superior to a no-instruction comparison and equally effective as a comprehensive control in which students received instruction for all conditional discriminations involved in the task. They also found that EBI students reach mastery criterion in less time and significantly fewer trials than students in the comprehensive instruction group do. Ong et al. (2018), found that new relations were evident only after EBI and that EBI was more effective and more efficient than both self-instruction and no-instruction controls.

In a study by Lovett et al. (2011), the effect of EBI was compared to a lecture format in teaching four single-subject designs to undergraduate students. Students who received EBI trained and tested baseline and emergent relations in a simple-to-complex (STC) protocol. Students in the lecture group viewed a 56 min video lecture with an accompanying PowerPoint presentation covering the same within-subject designs as arranged for EBI students. The results showed that a majority of the participants receiving EBI formed equivalence classes and that EBI was equally effective as the lecture-watching format. This study is particularly interesting because it compares the effectiveness of EBI compared to an often-used teaching format in higher education. However, concerns have been raised with regards to the effects of the traditional lecture format (Skinner, 1968), in which most of the time is allocated to the teacher talking and students listening. Within the field of behavior analysis, it has for a long time, been recognized that learning will not take place in the absence of reinforcing contingencies. When the teacher is lecturing and thereby playing the active part, the reinforcement schedule for the students' behavior will likely be thinner than if students actively do some kind of behavior. Traditional lectures is based upon the former alternative and therefore the probability of strengthening the desired behavior will decrease (Skinner, 1968). Keller (1968) proposes that students should participate in different learning activities for learning to be more effective.

Lately, and in the same vein, so-called student active learning has received increased attention and has gained popularity in higher education (e.g., Freeman et al., 2014; Martin et al., 2007). Student active learning is a collective term for a range of teaching formats characterized by an emphasis on in-class problem-solving assignments, discussions, worksheets or tutorials completed during class, use of personal response systems, peer instruction, and the minimal use of lectures. Student active learning has shown to be superior to lectures in promoting comprehension of the topics taught in undergraduate science, technology, engineering, and mathematics courses (Freeman et al., 2014). Thus, a comparison of the effectiveness and efficiency of EBI to a student active learning course would be an expansion of previous studies and may provide important information on the effects of EBI. Such a comparison could be done in a between-group design by providing EBI to one group of students and a student active learning format (SALF) to a second group and then comparing their performance on one or more dependent measures. On the other hand, EBI should not be considered the only teaching format in a course, but rather as one of several effective course components. Exposing students to both EBI and SALF conditions and controlling for the order of the interventions across groups, would allow comparison across conditions and could at the same time extend the findings in Lovett et al. (2011).

Lovett et al. (2011) arranged the EBI as a STC training and test protocol. In a STC protocol, symmetry is tested after the establishment of a single or a set of conditional discriminations and before the presentation of the conditional discrimination that will share one stimulus with the one previously established. Symmetry probes are repeated until the new relations reliably occurs. Then a new, or a set of new, conditional discriminations are introduced and trained separately to mastery criterion before symmetry are reliably demonstrated, followed by a test for transitivity and finally equivalence. The STC protocol has shown to generate high yields of equivalence class formation in basic studies (Adams et al., 1993; Fields et al., 1997). In a simultaneous protocol, on the other hand, all baseline relations are trained and established to criterion before a mixed test for all possible emergent relations. Since Lovett et al. (2011) arranged a STC protocol, an arrangement of a simultaneous protocol would have the potential to increase the generality of their study.

To sum up, equivalence procedures have proven to be an effective way to establish a variety of skills and academically relevant concepts in college students. However, the effectiveness and efficiency of EBI relative to other teaching formats are relatively limited.

An extension of the existing evidence in this line of research has the potential to provide valuable information about the usefulness of EBI and how to implement EBI in a course sequence. First, the present study aimed to do a systematic replication of Lovett et al. (2011) and to extend their investigation by comparing the relative effects of EBI to a SALF and with a simultaneous protocol. Second, we wanted to see whether the order in which EBI was introduced in a course would influence equivalence class formation, performance on the multiple-choice test, or scores on social validity questionnaires. Finally, we wanted to investigate whether EBI generated equivalence class formation and whether this would influence the outcome on a generalization test and a post-multiple-choice test.

Method

Participants

Eighty-one university students, 14 men, and 67 women between the age of 19 and 49 were recruited from two different classes. They all attended their first year of a bachelor's degree as social educators and were currently attending an introductory course on behavior analysis. At the onset of the experiment, 71 participants had participated in an introductory lecture on reinforcement, and 11 participants had applied experience with behavior analysis. Two participants reported having an earlier academic background within the behavior analytic field.

Before giving their consent, participants were told that the tasks were relevant for the upcoming exam, that participation was voluntary, and that they could withdraw at any time without any negative consequences. The same information was repeated at the onset of the EBI condition. When participants left the experimental room, they were thanked for participating and offered a full debriefing.

Setting and material

The experiment was conducted in a classroom setting. The matching-to-sample (MTS) training and testing were given in two classrooms with rows of four computers. Participants were seated 0.5–2 meters apart from one another. The computer cabinets were placed between stations to limit the view to the screens next to one another. One experimenter was seated in front of the classroom and another in the back.

HP70032 stationary computers (Microsoft Windows 10) were used to run the MTS training and testing. The computers had a 24-inch screen and an external mouse connected through a USB port. A custom-made software controlled stimulus presentation throughout the EBI condition. The program automatically registered and saved all important information such as trial type, correct/incorrect comparison choices, and programmed consequences given to the participant.

Design

We arranged a combined group and within-subject design, in which all participants experienced both the EBI and SALF conditions. One class was assigned to the EBI–SALF condition and the other to the SALF–EBI condition. The EBI–SALF group consisted of 33 participants (7 males and 26 females) between the ages 19–49 while the SALF–EBI group consisted of 48 participants (7 males and 41 females between the ages 19–39 years old). The mean age for both groups was 23 years old.

Figure 1 illustrates the arrangement of conditions for the two groups. The upper and lower panel shows the arrangement for the EBI–SALF and the SALF–EBI group, respectively. Participants in the EBI–SALF group experienced the EBI condition in an early phase of the course and then the SALF condition, while participants in the SALF–EBI group first experienced SALF and then EBI at the end of the course. Both groups experienced the multiple-choice test three times. At the onset of the course (T1), after SALF or EBI (T2), and finally at the end of the course (T3) three weeks later. The three multiple-choice tests were identical.

Dependent measures

Scores on the Multiple-Choice tests, T1, T2, and T3, were used to evaluate the effects of teaching format and order. Social validity scores were used as an additional measure for order effects. Finally, the tests for stimulus equivalence were used to assess the effect of the EBI condition. Percentage correct responding on generalization probes (AE and AF trials) and performance in the final multiple-choice test, T3, was compared for participants who responded in accordance with equivalence and those who did not.

Statistical analyses

A mixed ANOVA and post hoc (paired samples t-tests) were conducted to evaluate differences between instruction format and order.

Procedure

Multiple-choice tests

The multiple-choice tests were conducted as an online questionnaire. Appendix A provides examples of multiple-choice questions. Each example is listed with the

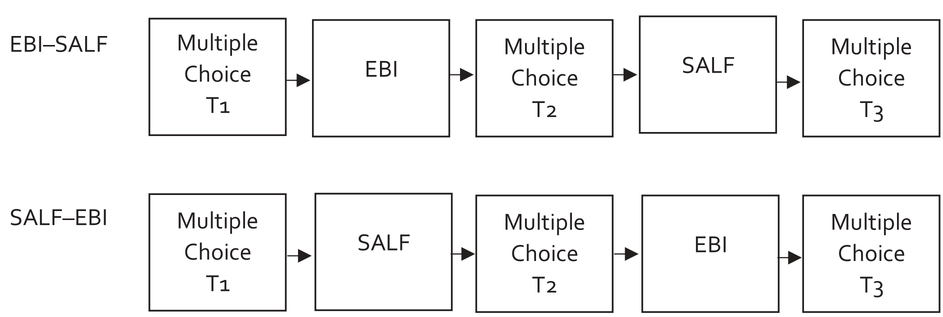

Figure 1. The figure is an illustration of the design and shows that the participants experienced the experimental conditions in two different orders. Some participants were introduced to MTS training and test for emergent responding (EBI) at an early stage of the course (upper panel), while other participants experienced student active learning format before matching-to-sample (lower panel).

corresponding alternatives below. Each multiple-choice test consisted of 14 questions with four response alternatives each. Nine questions targeted reinforcement, and five questions targeted punishment. The participants were not given any feedback on their choices. Questions addressed (1) whether a stimulus is added or removed, (2) identification of the corresponding procedure/process in different vignettes, (3) the procedural similarities and differences between reinforcement and punishment and positive and negative, and (4) pairing the statements about the concepts trained during the EBI condition to its corresponding definition.

Student active learning format

The class coordinator was informed about the topics included in the EBI condition but did not see the stimulus set used in this condition. A short lecture briefly introduced/presented each specific topic based on the curriculum, a Norwegian introduction book on behavior analysis (Horne & Oyen, 2005). Lectures typically lasted from 2 hours and 45 minutes to 3 hours and 15 minutes and consisted of traditional lecturing interspersed with peer tutoring, video clips demonstrating the principles, in-class discussions, short live demonstrations, answering oral questions (e.g., identify the correct procedure), and lecturer giving immediate feedback on tasks and the use of *Kahoot!*

The SALF condition was designed with elements from *flipped classroom* (Mok, 2014), where participants are required to read the relevant chapter before class attendance. Furthermore, the SALF condition included features from interteaching (Boyce & Hineline, 2002), in which participants were required to discuss the related study questions at the end of each chapter in groups of four to eight persons.

The duration of in-class discussion was up to two hours and was set by the class instructor. It typically included discussing 12–30 questions. Questions would review the content in the chapters the class coordinator and a teaching-assistant would take part in the discussions and help clarify misunderstandings when needed. At the end of each group discussion, participants were provided a pen and pencil test to fill in the blanks, which were ended with an in-class review of the correct answers. In total, there were six SALF sessions consisting of three introductory lectures on positive and negative reinforcement and positive and negative punishment, each followed by three group discussions. Participants spent approximately 15 hours in the SALF condition. Participants were required to attend the SALF sessions for at least 80% of the total time.

Equivalence-based instruction

For an overview of the arrangement of the EBI condition, see Figure 2. At the onset of the EBI condition, the experimenter first repeated general information about the experiment; that participation was voluntary, that they could withdraw at any time without any consequences, and that no personal information would be collected. Furthermore, they were instructed to turn off their cellular phones and not to talk to each other. Participants were allowed to take breaks outside the classroom at request. When participants had completed the EBI condition, they were offered a debriefing.

Stimuli. The stimulus set developed for this experiment was based on the course curriculum, a Norwegian introduction book on behavior analysis. Figure 3 provides an overview of the stimulus set used and shows that we arranged four potential classes,

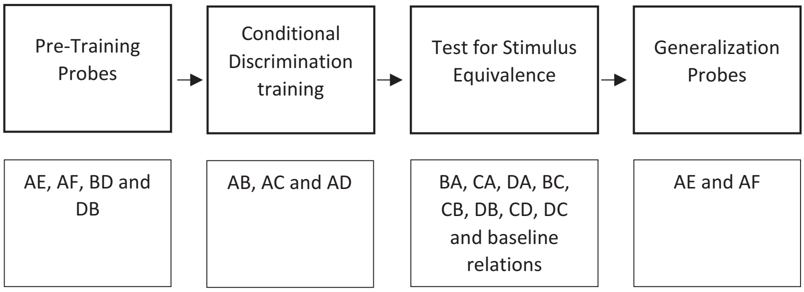

Figure 2. The figure shows the overall arrangement of the MTS training. The upper panel, where the boxes are connected with arrows shows the main phases of MTS training and test. The lower panel shows the trial types involved in each phase.

	1	2	3	4
a	POSITIVE REINFORCEMENT	NEGATIVE REINFORCEMENT	POSITIVE PUNISHMENT	NEGATIVE PUNISHMENT
b	STIMULUS IS PRESENTED DEPENDENT ON BEHAVIOR IN PRESENCE OF A GIVEN ANTECEDENT AND THE PROBABILITY OF THE BEHAVIOR INCREASES.	STIMULUS IS REMOVED DEPENDENT ON BEHAVIOR IN PRESENCE OF A GIVEN ANTECEDENT AND THE PROBABILITY OF THE BEHAVIOR INCREASES.	STIMULUS IS PRESENTED DEPENDENT ON THE BEHAVIOR IN PRESENCE OF A GIVEN ANTECEDENT AND THE PROBABILITY OF THE BEHAVIOR DECREASES.	STIMULUS IS REMOVED DEPENDENT ON THE BEHAVIOR IN PRESENCE OF A GIVEN ANTECEDENT AND THE PROBABILITY OF THE BEHAVIOR DECREASES.
c	ANTECEDENT: BEHAVIOR → REINFORCER ADDED	ANTECEDENT: BEHAVIOR → AVERSIVE REMOVED	ANTECEDENT: BEHAVIOR → AVERSIVE ADDED	ANTECEDENT: BEHAVIOR → REINFORCER REMOVED
d	I WONDER HOW A TERM IS DEFINED, READ THE CURRICULUM AND GET THE ANSWER. WHEN I LATER WONDER HOW A TERM IS DEFINED, I READ IN THE SAME BOOK TO FIGURE OUT.	I HAVE A MOSQUITO BITE AND SCRATCHES IT. THIS CEASES THE DISCOMFORT FOR A LITTLE WHILE. ON LATER OCCASIONS I SCRATCH AGAIN.	I DRINK MILK THAT IS EXPIRED AND GET NAUSEOUS. LATER, I ALWAYS CHECK THE EXPIRATION DATE AND NEVER DRINK MILK WHICH IS CLOSE TO THE EXPIRATION DATE.	I ALWAYS LEAVE THE POCKET IN MY JACKET OPEN UNTIL I ONE DAY LOSE MY PHONE AND THE SCREEN BREAKS. AFTER THIS I NEVER LEAVE THE POCKET OPEN.
e	A BEHAVIOR PRODUCES AN EVENT AND THE PROBABILITY OF THE BEHAVIOR IS INCREASED.	A BEHAVIOR REMOVES AN EVENT AND THE PROBABILITY OF THE BEHAVIOR IS INCREASED.	A BEHAVIOR PRODUCES AN EVENT AND THE PROBABILITY OF THE BEHAVIOR IS REDUCED.	A BEHAVIOR REMOVES AN EVENT AND THE PROBABILITY OF THE BEHAVIOR IS REDUCED.
f	I AM TIRED, EAT A PIECE OF CHOCOLATE AND GET MORE ENERGY. NEXT TIME I AM TIRED I DO THE SAME.	I GOT MIGRAINE AND EAT A PAIN KILLER. THE HEADACHE IS RELIEVED. NEXT TIME MY HEAD HURTS I GET A PAIN KILLER.	I GET A HEADACHE WHEN READING FOR A LONG TIME WITHOUT GLASSES. THE BEHAVIOR READING WITHOUT GLASSES DECREASES IN FREQUENCY.	I LOSE 100 NOK IN A BET WITH A FRIEND. AFTER THIS I NEVER BET AGAIN.

Figure 3. The figure displays stimuli used in the EBI condition. 1–4 represent the classes and A-F represent the members. A-D were trained and E and F were only used as a test for generalization.

corresponding with the concepts of positive reinforcement, negative reinforcement, positive punishment, and negative punishment. The figure is designed so thatcolumns 2–5 display the different concepts to be learned, and the rows show the various members of each concept. The different rows correspond with the name of the concepts to be formed

(A-stimuli), a written definition of each procedure (B-stimuli), short descriptions of the three-term contingency, specifying the consequence (C-stimuli), and written vignettes (D-stimuli). A, B, C, and D stimuli were included in MTS training and test for emergent relations. E and F stimuli were not included in the training, but presented as probes in pretest to test for pre-experimental history. E stimuli were novel definitions of the different concepts, containing the same information as the B stimuli, but described formulated differently, and F stimuli were novel vignettes.

Following the test for emergent relations, E and F stimuli were again presented as a test for generalization.

All stimuli were presented as black text on a white background. Since the stimuli varied with regards to the number of words, the size of the letters appeared slightly different for each stimulus. That is, stimuli with more letters appeared smaller while stimuli with fewer letters appeared a bit stretched, but still clear and readable. All stimuli had a click sensitive area of approximately 3–5 cm width and 4 cm height.

Instructions. At the onset of the experiment, the following instructions in Norwegian were projected in front of the classroom. Participants were asked to read it before starting the MTS program:

> When the experiment begins, a word or a sentence will appear on the screen. You should click on this. Alternative answers will then appear in the corners on the screen. You are to choose one of them. There will be no feedback in regards to correct or incorrect answers in the very first trials. After that, you will, on every trial, get feedback on whether your choice was correct or incorrect. When all tasks are learned, you will once again experience that you will not get feedback on your choices. Based on what you have already learned, it is still possible to get everything correct. Do your best to get everything correct. Good luck! Press Start to begin the experiment.

Pre-training probes. Pre-training probes were arranged to test for possible pre-experimental experience with the concepts. The following relations were included in this test: (a) concept name to definition II (AE relations), (b) concept name to vignette II (AF relations), (c) vignette I to definition (DB relations), and (d) definition to vignette I (BD relations). The different trial types were presented three times and in random order, constituting a total of 48 trials without programmed consequences.

MTS training. We arranged a simultaneous training protocol, in which all conditional discriminations were established to criterion before testing for emergent relations. Twelve baseline conditional discriminations were trained in a One-to-Many (OTM) training structure in which the A-stimuli served as sample stimuli. In the following, the different trial types are presented as letter and number combinations. The first letter and number combination refers to the sample stimulus and is separated from the four comparison stimuli by a slash. The correct comparison stimulus is underlined. The baseline trials were introduced in a serialized fashion, implying that all AB relations (A1/$\underline{\text{B1}}$B2B3B4, A2/B1$\underline{\text{B2}}$B3B4, A3/B1B2$\underline{\text{B3}}$B4, and A4/B1B2B3$\underline{\text{B4}}$) and AC relations (A1/$\underline{\text{C1}}$C2C3C4, A2/C1$\underline{\text{C2}}$C3C4, A3/C1C2$\underline{\text{C3}}$C4, and A4/C1C2C3$\underline{\text{C4}}$) were introduced separately and mastered to criterion, before the relations were presented in a mixed training block. Finally, the AD relations (A1/$\underline{\text{D1}}$D2D3D4, A2/D1$\underline{\text{D2}}$D3D4, A3/D1D2$\underline{\text{D3}}$D4, and A4/D1D2D3$\underline{\text{D4}}$) were introduced separately and mastered before participants experienced a mixed block of the AB, AC, and AD relations.

Each trial started with the presentation of a sample stimulus in the middle of the screen. A click to the sample was immediately followed by the presentation of four comparison stimuli in the corners of the screen. The location of the comparisons alternated across trials to prevent control by location. Following a comparison choice, a programmed consequence in blue letters on white background was presented in the middle of the screen for 500 ms. Correct comparison choices were followed by the Norwegian words for “good”, “perfect”, “correct,” and the like in the middle of the screen. Incorrect responses were followed by the Norwegian word for “wrong” presented in the middle of the screen. Programmed consequences were presented in blue letters on a white background. Each trial was followed by a 500 ms inter-trial interval. In each training block, the different trial types were presented four times each in random order. Training blocks were repeated until a criterion of 95% correct comparison choice was met.

Thinning of programmed consequences. Participants then experienced blocks of training in which programmed consequences were gradually thinned from 75% to 50% and finally 0% before testing for emergent relations. The criterion to proceed to the next training block was 95% correct comparison choice.

Test for stimulus equivalence. The test for stimulus equivalence presented all possible emergent relations and probes for maintenance of baseline relations. The following symmetry trials were tested for: B1/A1A2A3A4, B2/A1A2A3A4, B3/A1A2A3A4, B4/A1A2A3A4, C1/A1A2A3A4, C2/A1A2A3A4, C3/A1A2A3A4, C4/A1A2A3A4, D1/A1A2A3A4, D2/A1A2A3A4, D3/A1A2A3A4 and D4/A1A2A3A4. The symmetry trials were presented in a mixed block with equivalence trials. Equivalence trials were B1/C1C2C3C4, B2/C1C2C3C4, B3/C1C2C3C4, B4/C1C2C3C4, B1/D1D2D3D4, B2/D1D2D3D4, B3/D1D2D3D4, B4/D1D2D3D4, C1/B1B2B3B4, C2/B1B2B3B4, C3/B1B2B3B4, C4/B1B2B3B4, C1/D1D2D3D4, C2/D1D2D3D4, C3/D1D2D3D4, C4/D1D2D3D4, D1/C1C2C3C4, D2/C1C2C3C4, D3/C1C2C3C4 and D4/D1D2D3D4. Baseline trials were presented interspersed with emergent relations to test for maintenance. Each trial type was presented four times and in a random order, constituting a total of 240 trials. No programmed consequences were arranged during the test. The criterion to conclude that equivalence classes were formed, was a minimum of 90% correct comparison choices for baseline conditional discriminations, symmetry and equivalence respectively, and no more than one incorrect per trial type.

Probe for generalization. Immediately following the test for emergent relations, the participants underwent a probe test for AE and AF relations. Each relation was tested three times. This test was identical to the pre-test, except for any difference due to random presentation of trial types and that BD and DB trials were not included in the post-probes.

Social validity. Participants had no prior experience with EBI and a social validity questionnaire consisting of five questions was presented after the EBI condition to evaluate participants’ satisfaction with the new teaching format. They were asked to rate each question on a 7-point Likert scale. The questionnaire is shown in Appendix B.

Each question is listed along with the scale used. Higher scores indicate a more positive evaluation of the EBI condition. The questionnaire also included questions to assess participants' confidence in their own knowledge as well as time commitment.

Results

Participant flow

The flow of participants throughout the experiment is illustrated in Figure 4. The figure shows the EBI–SALF group to the left and the SALF–EBI group to the right. Boxes below each group display the different phases of the experiment, and the number of participants who started and fulfilled the different phases, in addition to reasons participants withdrew. The figure shows that 33 participants completed Multiple Choice T1 and started MTS training in the EBI–SALF group. Fifteen of the participants discontinued during MTS training. Two participants wrote notes during training and are therefore, not included in the data analysis. Finally, one participant was excluded due to a programming error. In total, fifteen of the participants carried out the MTS training as well as completed the Multiple Choice T2. Nine of them completed Multiple Choice T3. In the SALF–EBI group 48 participants completed Multiple Choice T1. Fourteen participants started MTS training.

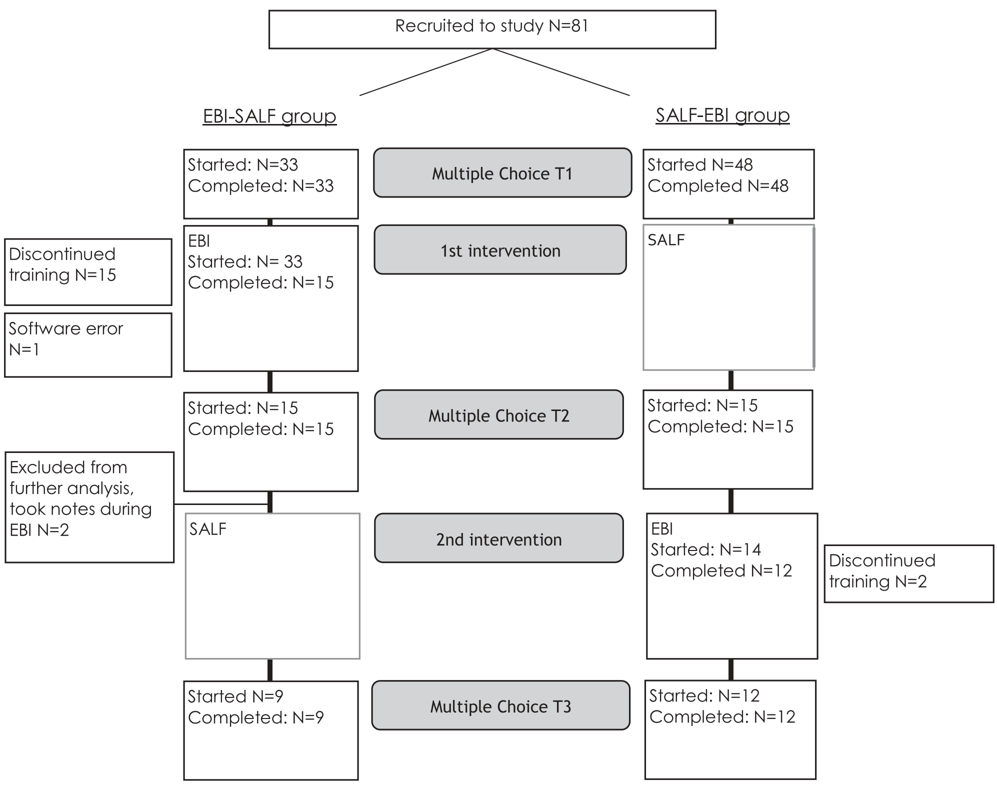

Figure 4. The figure shows a flow chart over the procedure, as well as how many participants completed each part in the two groups.

Two participants withdrew from the experiment while still in training, and 12 completed MTS training and test as well as Multiple Choice T2 and T3. Data from those who withdrew or were excluded from the study are not included.

Relative effects of the EBI and SALF conditions

Figure 5 shows the mean scores in the three multiple-choice tests for the two groups. Both groups showed slightly better performance at T2 and T3 compared to T1.

The mixed ANOVA revealed a significant main effect of teaching [$F(2, 40) = 14.387$, $p < 0.001$, $\eta^2_p = 0.418$]. Follow-up paired samples *t*-tests found a significant difference between T1 and T2 [$t(26) = 4.108$, $p < 0.001$], as well as between T1 to T3 [$t(21) = 5.373$, $p < 0.001$]. However, there was no significant difference between T2 to T3 [$t(21) = 1.267$, $p = 0.219$] leading to the conclusion that the two teaching formats were equally effective.

Participants spent 2–6 hours in the EBI condition, while in the SALF condition, participants spent approximately 15 hours.

Effects of order

The mixed ANOVA did not reveal any main effects of order [$F(1, 40) = 0.894$, $p = 0.356$ $\eta^2_p = 0.043$] nor interaction effects [$F(2, 40) = 1.148$, $p = 0.327$ $\eta^2_p = 0.054$]. However, social validity scores, indicating the participant's confidence of knowledge, can potentially help us get a broader understanding of order effects, as the test was run right after the EBI condition for both groups. Confidence of own knowledge of reinforcement was 3.9 for the EBI–SALF and 5.5 for the SALF–EBI group. For punishment, the scores were 3.8 and 5.5 for the groups, respectively.

On the social validity questionnaire, the EBI–SALF group scored the EBI condition as preferable to a mean of 2.9, as the SALF–EBI group scored it to 5.3. The time commitment was scored to 3.1 for the EBI-SALF and 5.1 for the SALF–EBI group. Finally,

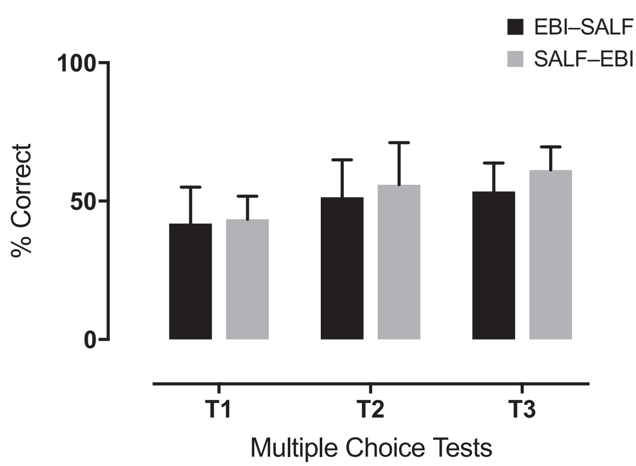

Figure 5. Scores on the Multiple-Choice Tests based from Group.
Note. The figure shows the mean score for the EBI–SALF group and the SALF–EBI group on the multiple-choice tests.

participants who experienced SALF before EBI scored the total duration of the training session to 1.5, while participants in the other group scored session length to 2.8.

Equivalence class formation

Table 1 shows the individual results for the stimulus equivalence test. The upper panel shows participants in the EBI–SALF group, while the bottom panel shows scores for the SALF–EBI group. The first column shows participant numbers, and the second column displays the number of trials above the minimum requirement. Column three–six show test performance as percentage correct for baseline, symmetry, and equivalence trials, respectively, and the last column shows trial types with more than one incorrect. Bolded numbers indicate scores in accordance with the equivalence criterion. In the EBI–SALF group, 11 out of 15 participants (73%) showed stimulus equivalence. Among participants who experienced the EBI condition at the end of the course, 9 out of 12 (75%) who fulfilled MTS training formed equivalence classes (Table 1). The mean training trials

Table 1. Individual scores at the test for stimulus equivalence.

	Training	Percent Correct Test			
Participant #	Trials >min.	BSL	SY	EQ	Trial Types with More than 1 Incorrect in Test
EBI–SALF					
18054	176	**96**	**100**	**100**	
18058	112	**100**	**100**	**98**	
18006	224	**96**	**100**	**97**	
18043	256	**98**	**100**	**97**	
18047	144	**100**	**98**	**99**	
18048	64	**96**	**98**	**98**	
18007	256	**98**	**98**	**98**	
18036	208	**100**	**98**	**97**	
18023	352	**96**	**98**	**96**	
18055	272	**100**	**98**	**96**	
18053	32	**90**	**96**	**99**	
18005	272	**100**	96	96	D2C2 and D3B3
18032	240	**100**	94	93	C1A1 and C4D4
18008	320	**96**	90	81	B2A2, B4C4, and C2B2
18057	160	71	79	70	A1C1, A2C2, A2D2, A3D3, B1C1, B2C2, B3A3, B4A4, C1D1, C3B3, C3D3, C4D4, D2C2, D3A3, and D4B4
SALF–EBI					
18022	64	**100**	**100**	**100**	
18083	96	**98**	**100**	**99**	
18091	48	**100**	**100**	**99**	
18070	144	**100**	**100**	**98**	
18092	80	**100**	**98**	**100**	
18081	16	**98**	**98**	**99**	
18073	112	**100**	**98**	**97**	
18082	16	**96**	**96**	**96**	
18071	48	**96**	**92**	**100**	
18029	560	**100**	**100**	95	C3D3
18069	128	**96**	**96**	89	D2B2
18080	144	**96**	88	**90**	B3C3

The upper part of the table shows participants who experienced equivalence-based instruction at the beginning of the course, and the lower part shows results in participants who experienced a student active learning format early in the course sequence. Trials >min denotes number of trials above the minimum. BSL denotes percentage of correct baseline trials under test, SY denotes the same for symmetry trials and EQ denotes equivalence trials. Trial types with more than one incorrect is specified to the right.

above the minimum for the EBI–SALF group was 205 and 121 for the SALF–EBI group. Participants 18029, 18032, and 18069 all responded 90% correct but erred more than once for one or more trial types. Participants 18008, 18057, 18080, and 18089 all had scores below the criterion of 90% correct for one or more relations. Regardless of order, participants made mistakes in both discriminating between negative and positive as well as reinforcement and punishment.

Figure 6 shows T3 results for participants who formed and did not form equivalence classes, regardless of the order of the instructional formats. Participants who met the criterion for stimulus equivalence had slightly better scores with a mean of 61% correct versus 50% correct for those who did not respond in accordance with stimulus equivalence. Participants who formed equivalence classes also show a better generalization to AE and AF probes. Figure 7 shows that mean scores on AE probes were 97% and 77%, and 89% versus 46% correct for AF probes for those who responded in accordance with stimulus equivalence and those who did not, respectively.

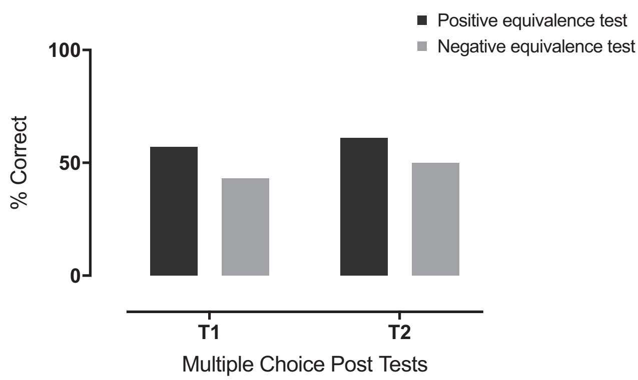

Figure 6. The figure shows percentage correct on the multiple-choice posttests. The black bars show participants with a positive equivalence outcome, while the grey bars show participants with a negative outcome on the stimulus equivalence test.

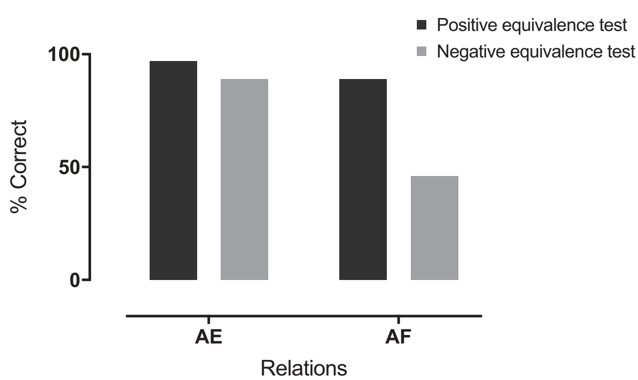

Figure 7. The figure shows percentage correct on the AE (new definition) and AF (new vignettes) trials during the post probe test. The black bars show participants with a positive equivalence outcome, while the grey bars show participants with a negative outcome on the test for stimulus equivalence.

Discussion

The study aimed to investigate the differential effects of two teaching formats and to illuminate whether the order in which the two formats were presented would influence the outcome on students' performance. Overall, the results showed an effect of both teaching formats on the multiple-choice tests, which replicate and expand the results in Lovett et al. (2011) in that they demonstrate that EBI can be as effective as other teaching formats. It also builds on existing literature in that EBI can promote social significant repertoires relevant for college students (e.g., Albright et al., 2015; Fields et al., 2009; Fienup & Critchfield, 2011; Fienup et al., 2010, 2016, 2015; Ong et al., 2018; Pytte & Fienup, 2012). By comparing to different teaching formats, the present study extended Lovett et al.'s (2011) findings. However, we found no significant effects related to the order of introducing the two formats, on performance in multiple-choice tests. Introducing the EBI condition at the end of the course resulted in fewer dropouts and better scores on the social validity questionnaire. Regardless of order, participants who formed equivalence classes also performed slightly better on two different tests for generalization.

Furthermore, the present experiment extended Lovett et al. (2011) by demonstrating that EBI was equally effective as SALF when EBI was arranged with a simultaneous protocol, which is associated with lower yields of equivalence than a STC protocol (Imam, 2006). However, results from studies are mixed. While some researchers (e.g., Adams et al., 1993; Fields et al., 1997) obtained high yields using a STC protocol in basic research, others (eg., Fienup & Critchfield, 2010; Fienup et al., 2010) obtained high yields of equivalence class formation using a simultaneous protocol in an applied setting. A comparison of the two in an applied setting (Fienup et al., 2015) replicates basic studies, which show better outcomes for the STC protocol.

Even though the outcome for the two conditions in the current study are equal, taking time spent in the different conditions into consideration, we must conclude that the EBI condition is less time consuming and, therefore, more efficient. Spending the limited class time on teaching formats that have proved to be effective is essential for students' learning outcomes. Like emphasized by Pytte and Fienup (2012), the efficiency of the equivalence paradigm opens for new possibilities in that more time can be spent going into depth of the material. The present results of learning basic concepts through EBI in a short amount of time, bear promise that more time can be allocated to other important issues such as why knowledge of these behavior principles is important and how they have been implemented to improve socially significant behavior of various kinds. Successful EBI presupposes considerable staff competence, technological equipment, training, and staff support. In sum, this may make the EBI less available for instructors. Eikeseth et al.

(1997) demonstrated that equivalence relations could be formed by using pen and paper. An extension of this study would be to provide the students with correct and incorrect answers, they would be able to self-administer the feedback. Furthermore, based on basic and applied research within the field of stimulus equivalence it is possible to extract which variables are most efficient in establishing the conditional discriminations and therefore, should be included in an applied experiment. Computer programs with default parameters can be made easily accessible for instructors without specialization within the field only to put in the stimulus set. This way it is possible to avoid some of the problems mentioned above.

The equivalence data from the present study show that over 70% of participants who went through training and testing reached the criterion for equivalence class formation. It is unlikely that participants within the same group discussed the relevant concepts with each other affecting the formation of stimulus equivalence classes as all training was done in one sitting. In the present study, we used a rigorous equivalence criterion. To conclude that an equivalence class was formed, the students would have to perform 90% correct on each of the three defining properties of stimulus equivalence and no more than one incorrect per trial type. This implies that participants can get 90% correct across the three properties but still not be evaluated to master the concepts. We would argue that a strict criterion is necessary to ensure that test performance reflects a satisfactory demonstration of the relevant concepts. If, for example, participants made two incorrect choices for one trial type, test performance for that trial type is random.

The MTS training format allows participants to use as many trials as necessary to establish baseline relations. The serialized trial arrangement implies that only four baseline relations are introduced in the same training block and that these are mastered before new trials are introduced. However, this is not sufficient for equivalence class formation in all participants. The participants who did not demonstrate emergent relations erred in both discriminating between positive and negative as well as reinforcement and punishment. To ensure that the necessary discriminations are established, one possibility worth looking into, in later experiments, is how training of the simple discriminations before the onset of training the conditional discriminations would affect the outcome on test for stimulus equivalence. Amongst others, the establishment of required prerequisites has been emphasized by several authors (e.g., McIlvane & Kledaras, 2012; Saunders & Spradlin, 1989) and may lead to a higher percentage of participants responding in accordance with equivalence. Nevertheless, for more than 70% of the participants, the training was sufficient to form stimulus equivalence. These findings are important as it shows that EBI can establish a basic understanding of concepts in a very precise way with a focus on how this can be done efficiently. However, when designing a course, it is important not only to consider which teaching formats to use but also the order in which the formats should be introduced.

The present results show that more participants fulfilled the EBI condition when first introduced to the concepts in sessions with SALF. Lovett et al. (2011) found that participants rate the EBI and lecture conditions as equally preferable. In the current study, participants only rated the EBI condition. The design makes it possible to make interpretations of the preference related to the order of presenting the EBI condition. Present results on the social validity questionnaire suggest that students might prefer EBI subsequent to SALF. It is reasonable to think that an introduction to the concepts given as SALF might give them an understanding of why the concepts are essential to learn and how the principles can be of applied value. The participants rate repeating tasks lower when they are not given any background information about why the concepts are important. The relatively low score for the EBI–SALF group might be, as also pointed out by Lovett et al. (2011), an effect of participants preferring what they are used to.

Overall, the EBI seems to be an efficient teaching format. It proves to be as effective as SALF, and participants use shorter time to establish the same repertoire as they do when given SALF. However, there are some limitations to the study. In the present experiment, participation was voluntary. This led to several participants in the EBI–SALF group did not conduct the last multiple-choice test. We do not know if or how this affects the outcome, but it must be considered as a limitation either way. One possible solution is to have attendance in experiments mandatory and still maintain participants' rights to withdraw from having their data used for research purposes.

Present results showed a small but statistically significant improvement from pretest (T1) to posttest (T2) for both experimental conditions, even though most participants showed stimulus equivalence. Thus, we did not see any effect of the order on the multiple-choice test, and the small difference from pre to post leads the attention to whether the difference is meaningful or not. Behavior analysts should aim towards robust and big differences in socially significant behaviors. Several features of the MC tests could cause a relatively small change. First, it can be difficult to draw a firm distinction between positive and negative reinforcement, even for experienced behavior analysts. For example, it may be difficult to differentiate whether the behavior of turning on the hot water in the shower is maintained by escape from cold water or adding warmth. The same may be the case in the vignette given as one of the questions in the multiple-choice test. Participants were asked to identify what might have happened in a scenario where parents pick up a crying child and the child starts to cry more often. Thus, participants might have reasoned correctly, but selected and incorrect alternative based on the premises given when scored by the experimenters. One solution would be to have participants provide a written rationale for their choice. A correct rationale can support that a participant has thought of the problem in the correct way. By including rationales for choices, the multiple-choice test would be more sensitive to changes. Second, the present MC test included relatively few vignettes targeting each concept. A possibility is to present the participants with several vignettes differing in complexity. In the example with the child being picked up, it is not possible to know whether the behavior is being positively or negatively reinforced. By providing several vignettes one can eliminate that one vignette is decisive. Last, one can specify the maintaining variable in the vignettes with several possibilities.

Furthermore, concept formation can be tested in several different ways. In the present study, it was used a selection-based format. Selection-based with a rationale is already mentioned. Topography-based were participants fill in the answers is another possibility. By measuring the behavior in different ways, one could avoid the possibility of a skewed definition of a concept affecting the outcome. Future studies should consider measuring the change in behavior in several different ways, such as fill-ins and multiple choice. One way of doing this would be to use data from the exam. In this experiment, the exam questions were not developed by the authors and therefore not included as a dependent measure.

However, the current study replicates earlier findings in that EBI is an effective method to establish a new concept repertoire in college students. Moreover, the EBI condition proved to be as effective as the SALF condition. Even though there are limitations linked to the multiple-choice tests, the findings bear promise to how EBI can be incorporated as part of a course. More research is needed to extend the knowledge of how post scores can be raised and how variables related to the order of introduction of the different teaching formats would influence the outcome. Despite these limitations, the current results contribute to the existing body of knowledge by comparing two different formats were participants actively take part in their learning.

Disclosure statement

The authors report no conflict of interest.

Ethical approval

The studies are performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its amendments. We do not collect personal data, and the Norwegian Centre for Research Data does not require preapproval for this kind of data.

Ethical Statement

The study is performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its amendments. We do not collect personal data, and the Norwegian Centre for Research Data does not require preapproval for this kind of data.

Informed consent

Informed consent was obtained from all participants included in the study

ORCID

Hanne Augland http://orcid.org/0000-0002-1620-1478
Erik Arntzen http://orcid.org/0000-0002-8471-1058

References

Adams, B. J., Fields, L., & Verhave, T. (1993). Effects of test order on intersubject variability during equivalence class formation. *The Psychological Record*, *43*(1), 133–152. https://doi.org/org/10.1007/BF03395907

Albright, L., Reeve, K. F., Reeve, S. A., & Kisamore, A. N. (2015). Teaching statistical variability with equivalence-based instruction. *Journal of Applied Behavior Analysis*, *48*(4), 883–894. https://doi.org/10.1002/jaba.249

Albright, L., Schnell, L., Reeve, K. F., & Sidener, T. M. (2016). Using stimulus equivalence based instruction to teach graduate students in applied behavior analysis to interpret operant functions of behavior. *Journal of Behavioral Education*, *25*(3), 290–309. https://doi.org/10.1007/s10864-016-9249-0

Arntzen, E., Halstadtro, L. B., Bjerke, E., & Halstadtro, M. (2010). Training and testing music skills in a boy with autism using a matching-to-sample format. *Behavioral Interventions*, *25*(2), 129–143. https://doi.org/10.1002/bin.301

Barron, R., Leslie, J. C., & Smyth, S. (2018). Teaching real-world categories using touchscreen equivalence-based instruction. *The Psychological Record*, *68*(1), 89 101. https://doi.org/10.1007/s40732-018-0277-0

Boyce, T. E., & Hineline, P. (2002). Interteaching: A strategy for enhancing the user friendliness of behavioral arrangements in the college classroom. *The Behavior Analyst*, *25*(2), 215–226. https://doi.org/10.1007/BF03392059

De Souza, D. G., de Rose, J. C., Faleiros, T. C., Bortoloti, R., Hanna, E. S., & McIlvane, W. J. (2009). Teaching generative reading via recombination of minimal textual units: A legacy of verbal behavior to children in Brazil. *International Journal of Psychology and Psychological Therapy*, *9*(1), 19–44.

Dixon, M. R., Belisle, J., Stanley, C. R., Daar, J. H., & Williams, L. A. (2016). Derived equivalence relations of geometry skills in students with autism: An application of the PEAK-E curriculum. *The Analysis of Verbal Behavior*, *32*(1), 38–45. https://doi.org/10.1007/s40616-016-0051-9

Eikeseth, S., Rosales-Ruiz, J., Duarte, A., & Baer, D. M. (1997). The quick development of equivalence classes in a paper-and-pencil format through written instructions. *The Psychological Record*, *47*(2), 275–284. https://doi.org/10.1007/BF03395225

Fields, L., Reeve, K. F., Rosen, D., Varelas, A., Adams, B. J., Belanich, J., & Hobbie, S. A. (1997). Using the simultaneous protocol to study equivalence class formation: The facilitating effects of nodal number and size of previously established equivalence classes. *Journal of the Experimental Analysis of Behavior*, *67*(3), 367–389. https://doi.org/10.1901/jeab.1997.67-367

Fields, L., Travis, R., Roy, D., Yadlovker, E., de Aguiar-rocha, L., Sturmey, P., & Ninness, C. (2009). Equivalence class formation: A method for teaching statistical interactions. *Journal of Applied Behavior Analysis*, *42*(3), 575–593. https://doi.org/org/10.1901/jaba.2009.42-575

Fienup, D. M., Covey, D. P., & Critchfield, T. S. (2010). Teaching brain–behavior relations economically with stimulus equivalence technology. *Journal of Applied Behavior Analysis*, *43*(1), 19–33. https://doi.org/10.1901/jaba.2010.43-19

Fienup, D. M., & Critchfield, T. S. (2010). Efficiently establishing concepts of inferential statistics and hypothesis decision making through contextually controlled equivalence classes. *Journal of Applied Behavior Analysis*, *43*(3), 437–462. https://doi.org/10.1901/jaba.2010.43-437

Fienup, D. M., & Critchfield, T. S. (2011). Transportability of equivalence-based programmed instruction: Efficacy and efficiency in a college classroom. *Journal of Applied Behavior Analysis*, *44*(3), 435–450. https://doi.org/10.1901/jaba.2011.44-435

Fienup, D. M., Mylan, S. E., Brodsky, J., & Pytte, C. (2016). From the laboratory to the classroom: The effects of equivalence-based instruction on neuroanatomy competencies. *Journal of Behavioral Education*, *25*(2), 143–165. https://doi.org/10.1007/s10864-015-9241-0

Fienup, D. M., Wright, N. A., & Fields, L. (2015). Optimizing equivalence-based instruction: Effects of training protocols on equivalence class formation. *Journal of Applied Behavior Analysis*, *48*(3), 613–631. https://doi.org/10.1002/jaba.234

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Green, G., & Saunders, R. R. (1998). Stimulus equivalence. In K. A. Lattal & M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 229–262). Springer.

Griffith, K. R., Ramos, A. L., Hill, K. E., & Miguel, C. F. (2018). Using equivalence-based instruction to teach piano skills to college students. *Journal of Applied Behavior Analysis*, *51*(2), 207–219. https://onlinelibrary.wiley.com/doi/pdf/10.1002/jaba.438

Hall, S. S., DeBernardis, G. M., & Reiss, A. L. (2006). The acquisition of stimulus equivalence in individuals with fragile X syndrome. *Journal of Intellectual Disability Research*, *50*(9), 643–651. https://doi.org/10.1111/j.1365-2788.2006.00814.x

Henklain, M. H. O., & Carmo, J. D. S. (2013). stimulus equivalence and increase of correct responses in addition and subtraction problems. *Paidéia (Ribeirão Preto)*, *23*(56), 349–358. https://doi.org/10.1590/1982-43272356201309

Horne, H. A., & Oyen, B. T. (2005). *Malrettet Miljoarbeid: Anvendt Atferdsanalyse. Del 1: Laeringsteori og Dagliglivets Pedagogikk*. GDR.

Imam, A. A. (2006). Experimental control of nodality via equal presentations of conditional discriminations in different equivalence protocols under speed and no-speed conditions. *Journal of the Experimental Analysis of Behavior*, *85*(1), 107–124. https://doi.org/10.1901/jeab.2006.58-04

Keller, F. S. (1968). Goodbye Teacher. *Journal of Applied Behavior Analysis*, *1*(1), 79–89. https://doi.org/10.1901/jaba.1968.1-79

Leader, G., & Barnes-Holmes, D. (2001). Establishing fraction-decimal equivalence using a respondent-type training procedure. *The Psychological Record*, *51*(1), 151–165. https://doi.org/10.1007%2FBF03395391

LeBlanc, L. A., Miguel, C. F., Cummings, A. R., Goldsmith, T. R., & Carr, J. E. (2003). The effects of three stimulus-equivalence testing conditions on emergent us geography relations of children diagnosed with autism. *Behavioral Interventions: Theory & Practice in Residential & Community-Based Clinical Programs*, *18*(4), 279 289. https://doi.org/10.1002/bin.144

Lovett, S., Rehfeldt, R. A., Garcia, Y., & Dunning, J. (2011). Comparison of a stimulus equivalence protocol and traditional lecture for teaching single-subjects designs. *Journal of Applied Behavior Analysis*, *44*(4), 819–833. https://doi.org/10.1901/jaba.2011.44-819

Lynch, D. C., & Cuvo, A. J. (1995). Stimulus equivalence instruction of fraction-decimal relations. *Journal of Applied Behavior Analysis*, *28*(2), 115–126. https://doi.org/10.1901/jaba.1995.28-115

Martin, T., Rivale, S. D., & Diller, K. R. (2007). Comparison of student learning in challenge-based and traditional instruction in biomedical engineering. *Annals of Biomedical Engineering*, *35*(8), 1312–1323. https://doi.org/10.1007/s10439-007-9297-7

McDonagh, E. C., McIlvane, W. J., & Stoddard, L. T. (1984). teaching coin equivalences via matching to sample. *Applied Research in Mental Retardation*, *5*(2), 177–197. https://doi.org/10.1016/S0270-3092(84)80001-6

McIlvane, W. J., & Kledaras, J. B. (2012). Some things we learned from Sidman and some we did not (We think). *European Journal of Behavior Analysis*, *13*(1), 97–109. https://doi.org/10.1080/15021149.2012.11434410

Mok, H. N. (2014). Teaching tip: The flipped classroom. *Journal of Information Systems Education*, *25*(1), 7–11.

Ninness, C., Barnes-Holmes, D., Rumph, R., McCuller, G., Ford, A. M., Payne, R., Ninness, S. K., Smith, R. J., Ward, T. A., & Elliott, M. P. (2006). Transformations of mathematical and stimulus functions. *Journal of Applied Behavior Analysis*, *39*(3), 299–321. https://doi.org/10.1901/jaba.2006.139-05

Ninness, C., Dixon, M., Barnes-Holmes, D., Rehfeldt, R. A., Rumph, R., McCuller, G., Holland, J., Smith, R., Ninness, S. K., & McGinty, J. (2009). Constructing and deriving reciprocal trigonometric relations: A functional analytic approach. *Journal of Applied Behavior Analysis*, *42*(2), 191–208. https://doi.org/10.1901/jaba.2009.42-191

Ong, T., Normand, M. P., & Schenk, M. J. (2018). Using equivalence-based instruction to teach college students to identify logical fallacies. *Behavioral Interventions*, *33*(2), 122–135. https://doi.org/10.1002/bin.1512

Pytte, C. L., & Fienup, D. M. (2012). Using equivalence-based instruction to increase efficiency in teaching neuroanatomy. *Journal of Undergraduate Neuroscience Education*, *10*(2), A125–A131.

Saunders, K. J., & Spradlin, J. E. (1989). Conditional discrimination in mentally retarded adults: The effect of training the component simple discriminations. *Journal of the Experimental Analysis of Behavior*, *52*(1), 1–12. https://doi.org/10.1901/jeab.1989.52-1

Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech, Language, and Hearing Research*, *14*(1), 5–13. https://doi.org/10.1044/jshr.1401.05

Sidman, M., & Cresson, O. (1973). Reading and crossmodal transfer of stimulus equivalences in severe retardation. *American Journal of Mental Deficiency*, *77*(5), 515–523. https://www.openathens.net

Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. Matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior*, *37*(1), 5–22. https://doi.org/10.1901/jeab.1982.37-5

Skinner, B. F. (1968). *The technology of teaching*. The B. F. Skinner Foundation.

Toussaint, K. A., & Tiger, J. H. (2010). Teaching early braille literacy skills within a stimulus equivalence paradigm to children with degenerative visual impairments. *Journal of Applied Behavior Analysis*, *43*(2), 181–194. https://doi.org/10.1901/jaba.2010.43-181

Appendices

Appendix A. Examples from multiple-choice test questions

Question
I get a headache when I am reading for a longer period without glasses. I am reading for a shorter and shorter period of time without first getting my glasses on. This is an example of; – Positive reinforcement – Negative reinforcement – Positive punishment – Negative punishment
Sophie gets angry when Emma grabs one of her toys Sophie hits Emma, which results in Emma giving Sophie's toy back. What can be said about this situation? – Sophie's problem behavior has been positively reinforced by getting the toy back – Sophie's problem behavior has been negatively reinforced by getting the toy back – Sophie's problem behavior has been positively punished by getting the toy back – Sophie's problem behavior has been negatively punished by getting the toy back
Sophie gets angry when Emma gets her toy. Sophie hitting Emma results in Emma not taking toys from Sophie. What maintains Emma's behavior? – The fact that Emma stops taking toys from Sophie is maintained by negative reinforcement – The fact that Emma stops taking toys from Sophie is maintained by positive reinforcement – The fact that Emma stops taking toys from Sophie is maintained by negative punishment – The fact that Emma stops taking toys from Sophie is maintained by positive punishment
The similarities between positive and negative punishment are – Both involve a stimulus change after the behavior has occurred – Both cause the behavior to decrease in frequency – Both options over are correct – None of the options over are correct

Appendix B. Social validity questionnaire

How confident do you feel in your knowledge of positive and negative reinforcement?						
1	2	3	4	5	6	7
Not at all			Somewhat confident			Very confident
How confident do you feel in your knowledge of positive and negative punishment?						
1	2	3	4	5	6	7
Not at all			Somewhat confident			Very confident
Rate the degree to which you would prefer to be taught using this instructional method						
1	2	3	4	5	6	7
Don't prefer			OK			Strongly prefer
How appropriate was the time commitment for this instructional method in relation to the amount you feel you have learned?						
1	2	3	4	5	6	7
Not at all			Somewhat appropriate			Very appropriate
How do you feel about the length of this instructional method?						
1	2	3	4	5	6	7
Too long			OK			Too short