Educational Relevance as a Motivating Operation for Equivalence Class Formation: Implications for Application

Julia Brodsky
The Graduate Center, City University of New York
EDUCATIONAL RELEVANCE AS A MOTIVATING OPERATION FOR EQUIVALENCE
CLASS FORMATION:

IMPLICATIONS FOR APPLICATION

by

JULIA BRODSKY

A dissertation submitted to the Graduate Faculty in Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

2018
Educational Relevance as a Motivating Operation for Equivalence Class Formation:
Implications for Application

by

Julia Brodsky

This manuscript has been read and accepted for the Graduate Faculty in Psychology in
satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Date

Robert Lanson
Chair of Examining Committee

Date

Richard Bodnar
Executive Officer

Supervisory Committee:

Daniel Fienup

Alexandra Logue

Kenneth Reeve

Janet Twyman

THE CITY UNIVERSITY OF NEW YORK
ABSTRACT

Educational Relevance as a Motivating Operation for Equivalence Class Formation: Implications for Application

by

Julia Brodsky

Advisor: Robert Lanson

Equivalence-Based Instruction (EBI) is a method of teaching sets of physically disparate stimuli that uses stimulus equivalence principles (Critchfield & Fienup, 2008). The basic stimulus equivalence literature informs which procedural variants are used in research and clinical practice, but application of these findings requires determining the effect of educational relevance on equivalence class formation. In the current study, Experiment 1 compared the relative difficulty of the stimulus sets to be used in the educationally relevant conditions of Experiment 2, while Experiment 2 examined the effects of stimulus meaningfulness and educational relevance. The Arbitrary group learned nonsense symbols from Steele and Hayes (1991). The meaningful/ non-educationally relevant group (Chem) learned stimuli adapted from two Chemistry textbooks. The meaningful/ educationally relevant group (DevPsych) learned stimuli adapted from a Psychology textbook. Experiment 2 compared equivalence class formation in Arbitrary, Chem, and DevPsych groups, demonstrating that meaningfulness facilitates equivalence class formation relative to arbitrariness, and that educationally relevant stimuli facilitate equivalence class formation relative to non-educationally relevant stimuli. Experiment 1 demonstrated that the effect of educational relevance found in Experiment 2 was not due to stimulus difficulty alone. Together, findings from the current study show that the effect of educational relevance as a motivating operation may improve learning outcomes in EBI, explaining why procedural variants of EBI shown in the literature to be less effective may still produce high yields (i.e., percentage of participants passing the first administration of the final
test of equivalence) when used in applied contexts. This finding calls for a re-investigation of
basic stimulus equivalence findings in the applied setting.

*Keywords*: stimulus equivalence, equivalence-based instruction, concept formation,
meaningfulness, educational relevance
ACKNOWLEDGEMENTS

There are a number of people without whom this document would not have been possible. First, I would like to thank Dr. Robert Allan for introducing me to this amazing field and for giving me the opportunity to work on a number of basic research projects as an undergraduate student. I owe a tremendous thank you to Dr. Daniel Fienup, who has been an immensely supportive mentor throughout my time as a graduate student, gave me the opportunity to work on a broad range of translational and applied projects, and has always been readily available whenever I’ve had research or professional issues. I could not have completed such a tremendous project without your guidance. Thank you to Dr. Robert Lanson, who gave me the opportunity to resume basic research and dedicated so much time to co-mentoring this dissertation using his expertise in stimulus and experimental control. I really appreciate your patience and meticulous manuscript edits and comments. Thank you to Dr. Lexa Logue for her incredibly thoughtful feedback and support as an internal committee member. Thank you also to my external committee members, Dr. Kenneth Reeve and Dr. Janet Twyman, who provided thought-provoking comments and careful feedback on this manuscript. Thank you to Dr. Jennifer Stewart, who guided me in choosing the correct statistical analyses for this project. A big thank you to Alan Brice, whose unending support and patience encouraged me to preserve throughout this process. Thank you to my family and friends who provided me with social support throughout graduate school. Thank you to my colleagues for being so collaborative, having interesting discussions about behavior analysis at all hours of the day, and for always making me laugh. Last but not least, thank you to Danielle Lucania for keeping me on track with procedures, forms, and policies throughout my time in the program.
Table of Contents

Introduction ........................................................................................................................................... 1

From Applied Beginnings to the Basic Lab and Back ........................................................................... 2

Outcome Discrepancies .......................................................................................................................... 3

Potential Variables Impeding Translation ............................................................................................ 5

Meaningfulness ....................................................................................................................................... 5

Educational Relevance .......................................................................................................................... 8

The Current Study .................................................................................................................................. 9

Experiment 1 ........................................................................................................................................... 10

Method .................................................................................................................................................. 11

Participants ........................................................................................................................................... 11

Setting .................................................................................................................................................. 11

Stimuli, Software, and Materials .......................................................................................................... 11

Design .................................................................................................................................................. 13

Procedure .............................................................................................................................................. 13

Results and Discussion ......................................................................................................................... 16

Dependent Measures ............................................................................................................................ 16

Pretest Responding ................................................................................................................................. 18

Effectiveness ......................................................................................................................................... 19
Efficiency ......................................................................................................................... 19
Social Validity .................................................................................................................... 19
Summary ............................................................................................................................ 20

Experiment 2 .................................................................................................................... 21
Method ............................................................................................................................... 22

Participants and Setting ..................................................................................................... 22
Stimuli, Software, and Materials ....................................................................................... 22
Design ................................................................................................................................. 22

Procedure .......................................................................................................................... 22

DevPsych MO Manipulation .............................................................................................. 23
Chem MO Manipulation ...................................................................................................... 24
Arbitrary MO Manipulation ............................................................................................... 25

Results and Discussion .................................................................................................... 26

Dependent Measures ........................................................................................................ 26
Pretest Responding ........................................................................................................... 26

Effect of Meaningfulness .................................................................................................. 27

Effectiveness ..................................................................................................................... 27

Efficiency .......................................................................................................................... 27

Effect of Educational Relevance ....................................................................................... 28
Effectiveness .......................................................................................................................... 28

Efficiency ........................................................................................................................... 28

Social Validity .................................................................................................................... 29

Summary ............................................................................................................................. 30

General Discussion .......................................................................................................... 30

Appendices ......................................................................................................................... 62

References .......................................................................................................................... 65
LIST OF TABLES

Table 1. Stages and Steps of Conditional Discrimination Training...............................37

Table 2. Descriptive Statistics for Experiment 1 Tutorial Variables........................................38

Table 3. Descriptive Statistics for Experiment 1 Survey Responses .......................................39

Table 4. Descriptive Statistics for Experiment 2 Tutorial Variables .....................................40

Table 5. Descriptive Statistics for Experiment 2 Survey Responses .....................................41
LIST OF FIGURES

Figure 1. Stimuli used in the Chem condition ..........................................................42

Figure 2. Stimuli used in the DevPsych condition ..................................................43

Figure 3. Procedure used in Experiment 1 ..............................................................44

Figure 4. Steps in the fading procedure ..................................................................45

Figure 5. The two trial types presented in Stage 3 of fading and Stage 1 of component conditional discrimination training .........................................................46

Figure 6. The three trial types presented in Stage 5 of fading training ......................47

Figure 7. The two trial types presented in Stage 2 of component training ..................48

Figure 8. The four trial types presented in Stage 7 of fading and Stage 3 of component training 49

Figure 9. Scores on the 3Mix pretest for the Chem and DevPsych conditions in Experiment 1 50

Figure 10. Scores on the 3Mix posttest for the Chem and DevPsych conditions in Experiment 1 ........................................................................................................51

Figure 11. Percentage of participants reaching mastery criterion on the first administration of the 3Mix posttest in the Chem and DevPsych conditions in Experiment 1 .................52

Figure 12. Number of training trials and percentage of training errors for the Chem and DevPsych conditions in Experiment 1 .................................................................53

Figure 13. Cycle time for the Chem and DevPsych conditions in Experiment 1 ............54

Figure 14. Stimuli used in the Arbitrary condition ....................................................55

Figure 15. Procedure used in Experiment 2 ............................................................56
Figure 16. Scores on the 3Mix pretest for the Arbitrary, Chem, and DevPsych conditions in Experiment 2 .................................................................57

Figure 17. Scores on the 3Mix posttest for the Arbitrary, Chem, and DevPsych conditions in Experiment 2 ........................................................................58

Figure 18. Number of training trials and percentage of train errors for the Arbitrary, Chem, and DevPsych conditions in Experiment 2 ........................................................................59

Figure 19. Cycle time for the Arbitrary, Chem, and DevPsych conditions in Experiment 2 ........60

Figure 20. Percentage of participants reaching mastery criterion on the first administration of the 3Mix posttest in the Chem and DevPsych conditions in Experiment 2 .........................61
Educational Relevance as a Motivating Operation for Equivalence Class Formation: Implications for Application

Equivalence-Based Instruction (EBI) is an instructional design tool based on the stimulus equivalence paradigm. Using EBI, instructors teach at least two overlapping conditional discriminations to mastery, and additional untaught relations emerge during testing. The benefit of EBI is that it is generative and economical (Critchfield & Fienup, 2008). EBI’s ability to produce efficient learning gives this technology the potential to help college students succeed in their courses while minimizing instructional resources, thus allowing students to allocate their time to additional educational activities. Instructors can use this technology to design lecture content (e.g., Pytte & Fienup, 2012), develop homework assignments, or to supplement in-class instruction (e.g., Fienup, Mylan, Brodsky, & Pytte, 2016).

EBI is typically conducted with match-to-sample (MTS) procedures. A sample (question) and two or more comparisons (answers) are presented to a participant. Two or more overlapping conditional discriminations are trained, and a number of untaught relations emerge (Critchfield & Fienup, 2008). For example, Fields et al. (2009) trained four 4-member (A, B, C, D) classes of statistical interactions to participants. Within these classes, graphs of the interactions served as A stimuli, descriptions of graphs served as B stimuli, names of interactions served as C stimuli, and definitions of the interactions served as D stimuli. First, three baseline relations (A → B, B → C, and C → D) were trained to mastery. Next, the researchers tested for the emergence of three types of relations: Symmetry, transitivity, and equivalence. A participant demonstrated symmetry when A was selected in the presence of B (B → A), B in the presence of C (C → B), and

---

¹ In the equivalence literature, it is common practice to notate relations such that the letter representing the sample stimulus is written first, followed by an arrow, and then the letter representing the comparison stimulus is written last (e.g., A → B).
C in the presence of D (D→C). A participant demonstrated transitivity when s/he selected C in the presence of A (A→C), D in the presence of B (B→D), and D in the presence of A (A→D). The participant demonstrated equivalence, the combination of symmetry and transitivity, when s/he selected A in the presence of C (C→A), A in the presence of D (D→A), and B in the presence of D (D→B). Once a participant had demonstrated all three properties, s/he had formed a statistical interaction equivalence class whereby each stimulus evoked the selection of all other members of the class. Researchers determine the success of an equivalence procedure by measuring the effectiveness and efficiency of equivalence class formation. Effectiveness may be measured by variables such as posttest scores and the percentage of participants passing the first administration of the test of equivalence class formation. Efficiency may be measured by variables such as training time and trials, time to equivalence class formation, and percentage of training errors.

**From Applied Beginnings to the Basic Lab and Back**

Although researchers conducted initial equivalence work in applied settings (Sidman, 1971), much EBI research has been conducted in basic and translational settings to determine the underlying properties and generality of this technology (Fienup & Brodsky, *in press*). Basic stimulus equivalence researchers explore the variables affecting equivalence class formation in a laboratory setting. Doing so allows researchers to control for a variety of relevant variables, demonstrates strong internal validity, and gives researchers more confidence that observed effects are due to the manipulated variable rather than extraneous variables. One such relevant variable is reinforcement history for responding to certain stimuli. To control for reinforcement history, researchers teach classes that contain symbols or unfamiliar Greek, Arabic, and Hebrew letters (e.g., Arntzen, 2004).
Translational researchers study equivalence somewhere in the middle of the basic-applied continuum, potentially teaching socially significant concepts to participants for whom the content is not relevant, usually in a laboratory setting. Translational researchers often investigate parameters of equivalence (e.g., training protocol, or order of training and testing trials; Fienup, Wright, & Fields, 2015) in a controlled setting with content that may benefit a broader population. Such an example of a translational study was conducted by Trucil, Vladescu, Reeve, DeBar, and Schnell (2015) who taught portion-size estimation to students enrolled in a non-related degree program.

Applied researchers conduct studies in naturalistic settings, where EBI can teach learners repertoires that are socially significant to them. Although researchers must sacrifice control over relevant variables to demonstrate the generality of this technology, applied research is beneficial because it concerns educational uses that can directly benefit learners and their communities. In an example of an applied study, Lovett, Rehfeldt, Garcia, and Dunning (2011) used EBI to teach students enrolled in a Research Methods course about single-subject research designs. After completing instruction, participants demonstrated class-consistent responding to the names, definitions, graphs, and vignettes of four types of single-subject research designs.

**Outcome Discrepancies**

Both basic and translational research have revealed relatively more effective training and testing parameters for EBI. For example, research on training structure, or the arrangement of overlapping conditional discriminations, has shown that the many-to-one (MTO; B→A and C→A) and one-to-many (OTM; A→B and A→C) training structures produce higher yields (i.e., percentage of participants passing the first administration of the final test of equivalence) than the linear series (LS; A→B and B→C) training structure (Arntzen & Holth, 1997). Despite this
finding in highly controlled settings using arbitrary stimuli, many applied studies have used the LS training structure (i.e., teaching baseline conditional discriminations in sequence with multiple linking stimuli; Saunders & Green, 1999) with success (e.g., Fields et al., 2009; McGinty et al., 2012; Ninness et al., 2006). Further, Fienup et al. (2015) conducted a translational study comparing two common EBI training protocols (i.e., the manner in which training and testing phases are arranged) and discovered that intermixing training and testing phases, or the Simple-to-Complex training protocol (STC), improved equivalence class yields relative to completing all training phases followed by one testing phase (Serial Simultaneous training protocol; SIM).

However, Fienup et al. (2015), which was conducted in a laboratory setting with applied stimuli, found that the SIM protocol produced substantially higher outcomes (42% yield) than when the SIM training protocol has been evaluated in basic research (16% yield in Fields, Landon-Jimenez, Buffington, & Adams, 1995). Furthermore, a number of studies conducted in educational settings with applied stimuli have used the SIM training protocol with success (e.g., Pytte & Fienup, 2012; Sella, Ribeiro, & White, 2014). Additionally, basic research on MTS procedures demonstrates that the delayed MTS procedure facilitates equivalence class formation relative to simultaneous MTS procedures; yet most translational and applied equivalence studies use simultaneous MTS with success (e.g., Fienup et al., 2015). Therefore, previous research indicates that the level of success of a particular training variant (e.g., training protocol) depends upon whether the research context is basic, translational, or applied, which can either attenuate differences between treatment effects or enhance the treatment effect (e.g., training sequence). Despite basic and translational findings demonstrating that particular procedures work better than others, these findings may not necessarily have implications for the manner in which
research should be conducted in an applied setting, because participant performance does not necessarily suffer when these less effective procedural variations are used (e.g., SIM protocol, Sella et al., 2014; simultaneous MTS, Fienup et al., 2015).

**Potential Variables Impeding Translation**

A number of variables differ between basic stimulus equivalence research and translational and applied EBI research. These include the study context, which varies from tightly controlled laboratory settings (Mensah & Arntzen, 2016) to educational settings (Pytte & Fienup, 2012), and the type of stimuli in potential classes. Recently, researchers have begun investigating the type of stimuli and their influence on equivalence class formation.

**Meaningfulness.** Meaningfulness refers to a participant’s reinforcement history with a stimulus. Because the differences in the effects of basic, translational, and applied research paradigms depend heavily on a participant’s reinforcement history with the EBI content, it is important for researchers to determine the degree to which reinforcement history with stimuli impacts equivalence class formation. Arntzen (2004) studied the impact of stimulus reinforcement history on equivalence class formation by training three 5-member classes. Researchers assigned 50 university students to one of four conditions. In Condition 1, all stimuli presented were Greek and Arabic letters. In Condition 2, stimuli B through E were Greek letters, but stimulus A was a picture. In Condition 3, stimuli A through D were Greek letters, but stimulus E was a picture. In Condition 4, stimuli B through E were Greek letters, but the A stimulus was a nonsense syllable. The percentages of participants passing the first administration of the final test of equivalence (i.e., yield) for the four conditions was as follows: 30% for condition 1, 100% for condition 2, 50% for condition 3, and 40% for condition 4. Because picture placement as the A or E stimulus also varied, it is not entirely clear whether yield
differences were solely due to stimulus type; but yield differences between Conditions 1 and 2 suggest that meaningful stimuli (e.g., pictures that an individual is very likely to have experience with) may facilitate equivalence class formation relative to arbitrary stimuli (e.g., nonsense syllables, unfamiliar letters).

Fields, Arntzen, Nartey, and Eilifsen (2012) replicated and extended the work of Arntzen (2004) by manipulating participants’ reinforcement history with the training stimuli. Researchers taught one of three 5-member classes to each of 30 college students. Participants in the abstract (ABS) condition were exposed to abstract stimuli, which included a combination of Japanese, Greek, and Hebrew letters. Training stimuli for the picture (PIC) condition differed in that the C stimuli were replaced with pictures of a church, crown, and mailbox. Participants in the acquired (ACQ) condition experienced discrimination training that established their abstract C stimuli as discriminative stimuli. Results showed that the PIC condition produced the highest yield (80%), while the ABS condition produced 0% yield. The ACQ condition produced 50% yield, suggesting that the degree of reinforcement history a participant has with training stimuli affects equivalence class formation.

Mensah and Arntzen (2016) studied meaningfulness in a novel way to determine whether there was a functional relation between the number of classes containing meaningful stimuli and equivalence class formation. Sixty college students learned three 5-member classes. The 0PIC condition exposed participants to arbitrary stimuli only, which included a combination of Japanese, Greek, and Hebrew letters. The 1PIC condition exposed participants to one class with a meaningful stimulus as the C stimulus, a picture of a church. The 2PIC condition exposed participants to two classes with a meaningful stimulus (church and crown) and the 3PIC condition exposed participants to three classes, each with a meaningful stimulus (church, crown,
In all conditions, pictures were only used as a C stimulus to control for the effects of placement on equivalence class formation. Results showed that increasing the number of classes with a familiar C stimulus enhanced equivalence class formation, such that the 0PIC and 1PIC conditions resulted in 13.3% yield, the 2PIC condition resulted in 53.3% yield, and the 3PIC condition resulted in 80% yield. Results from this study demonstrated that the number of classes including meaningful stimuli positively impacts equivalence class formation.

Findings from Arntzen (2004), Fields et al. (2012), and Mensah and Arntzen (2016) demonstrate that meaningfulness of stimuli facilitates equivalence class formation relative to arbitrariness, and increased numbers of meaningful stimuli in a stimulus set increases the probability of equivalence class formation. These findings pose an issue for studying other variables that affect the formation of meaningful equivalence classes because the effect of meaningful stimuli may obfuscate efficiency and effectiveness differences between experimental manipulations in an applied research context that are apparent in the basic research context. For example, Arntzen and Holth (1997) showed that with volunteer participants learning arbitrary content, the LS training structure produces the lowest yields when compared to MTO and OTM; however the LS training structure has produced larger yields in the applied context, such as with Behavioral Neuroscience students learning neuroanatomy stimuli (i.e., “meaningful” stimuli; Fienup et al., 2016). Fienup et al. (2016) examined the effect of temporal ordering of stimuli (e.g., training name to picture then picture to function vs. picture to name then name to function) using the LS training structure, but researchers did not observe low yields because the effect of meaningfulness may have been stronger than the temporal ordering of stimuli.

**Educational Relevance.** It is also useful to investigate whether the context in which equivalence classes are trained affects equivalence class formation. The current study considers
the context in which instruction takes place. The context determines the degree to which the training stimuli are educationally relevant to the individual. This continuum spans from the stimuli not being educationally relevant to the learner, as is the case with abstract stimuli, to possibly relevant in the future or immediately educationally relevant, as may be the case with meaningful stimuli.

The educational relevance of the stimuli to the learner may serve as a motivating operation (MO; Cooper, Heron, & Heward, 2007). The behavior-altering effect of the MO can either be evocative, in that it increases the frequency of a behavior that is followed by the consequence, or abative, in that it decreases the frequency of a behavior that is followed by the consequence. The behavior-altering effect of the MO is a result of the change in reinforcer or punisher potency (Laraway, Srycerski, Michael, & Poling, 2003). That is, because the potency of a consequence has increased or decreased, the behavior that the reinforcer has previously followed will in turn also increase or decrease, respectively. Stimuli that are educationally relevant to the learner may serve as an establishing operation (EO)—that is, they increase the momentary effectiveness of the associated consequence for learning educationally relevant stimuli. In turn, hearing a sound associated with correct responding and advancing a counter that tracks progress toward a mastery criterion may increase in reinforcer potency. Hearing a sound associated with an incorrect answer or resetting a mastery criterion counter may increase punisher potency. In turn, the learner may attend to stimuli more carefully, thereby committing fewer errors, and scoring higher on the final test of equivalence. Stimuli that are not educationally relevant to the learner may serve as an abolishing operation (AO)—that is, they decrease the momentary effectiveness of the associated consequence for learning the non-educationally relevant stimuli. In turn, hearing a sound associated with correct responding and
advancing the mastery criterion counter decrease in potency as reinforcers. Hearing a sound associated with an incorrect answer or resetting a mastery criterion counter may decrease in potency as punishers. The learner may then attend less to stimuli, leading to an increase in trials, learning time and errors, and lower scores on the final test of equivalence.

To date, researchers have not studied the degree to which the educational relevance of stimuli in to-be-formed classes impacts equivalence class formation, only whether the stimuli are abstract or contain up to one meaningful stimulus per class. There is clear evidence from translational (Sella et al., 2014) and applied studies (Critchfield, 2014) that EBI is effective in teaching concepts to participants. The basic literature shows that particular variations (e.g., training structure; Arntzen & Holth, 1997) of equivalence procedures are effective in teaching concepts. Applied college-level studies generally explore the use of EBI in novel content areas, such as music (Hayes, Thompson, & Hayes, 1989) and mathematical functions (Ninness et al., 2006), or compare EBI to typical classroom instruction such as lecture (Lovett et al., 2011) and reading textbook passages (O’Neill et al., 2015). For researchers to uncover the variables that influence the formation of educationally relevant equivalence classes, it is important to determine whether using educationally relevant stimuli as an EO for learning increases scores on effectiveness and efficiency measures, and in what contexts the study of EBI is likely to be more informative.

The Current Study

The purpose of the current study was to determine whether there is an effect of meaningfulness and educational relevance in the formation of equivalence classes. Experiment 1 was designed to answer the following experimental question: Will the meaningful stimuli chosen for Experiment 2 produce similar baseline performances, in the absence of a MO manipulation,
using an efficient training procedure? The primary manipulations of meaningfulness and educational relevance were conducted in Experiment 2. Experiment 2 was designed to answer the following experimental questions: 1) is there an effect of stimulus meaningfulness on equivalence class formation? and 2) is there an effect of educational relevance on equivalence class formation? The Arbitrary group learned nonsense stimuli drawn from Steele and Hayes (1991). The two meaningful groups differed along the educational relevance continuum. The meaningful/ non-educationally relevant group, Chem, learned organic chemistry stimuli adapted from Baker, Rizzo, and Engel (2010) and Rittner and Bailey (2005). The meaningful/ educationally relevant group, DevPsych, learned developmental psychology stimuli adapted from Morris and Maisto (2005). MO manipulations, in the form of a series of statements and corresponding questions about those statements, were used in an attempt to increase the momentary reinforcing value of learning stimuli presented in the DevPsych condition, and decrease the momentary reinforcing value of learning Chem stimuli.

**Experiment 1**

Experiment 1 assessed baseline and acquisition similarities between the stimulus sets chosen for the educationally relevant and non-educationally relevant conditions. In order to use these two conditions as the meaningful conditions in Experiment 2, the DevPsych and Chem stimulus sets had to be either of equal difficulty or the Chem stimulus set had to produce more effective and efficient equivalence class formation outcomes. Such a finding would help rule out stimulus difficulty (e.g., more training trials, more training errors, more cycle time) alone as an explanation for any outcomes in Experiment 2 that may have indicated that the DevPsych condition produces more effective or efficient outcomes.
Method

Participants. Twenty-seven General Psychology students at different stages of their education and of the course participated for course credit. Participants gave informed consent to participate in a two-hour experiment. Demographic information for the participant pool from which students were drawn were as follows: 90% were between 18 and 25 years of age, 70% were born in the United States, 37% were Freshmen, 63% were male, 64% spoke English as their native language, and 28% were white. Such data are not available for the sample, but the following demographic academic data were collected from the demographics questionnaire (see Appendix B): 30% of participants planned to major in Psychology, 0% planned to major in Chemistry, 15% of participants had had experience working in a research lab, 15% of participants were interested in pursuing a research career, 11% were interested in a graduate degree in Chemistry or a related field, and 19% were interested in pursuing a career in healthcare.

Setting. All sessions took place in a 2.73 m by 2.55 m room containing three 1.35 m by 0.71 m cubicles. Each cubicle contained a Dell Optiplex© 760 desktop computer that used a Windows 7© operating system, a 0.40 m by 0.29 m monitor, a keyboard, a mouse, a desk, a chair, and headphones. Participants viewed stimuli displayed on a computer monitor and responded using the mouse.

Stimuli, Software, and Materials. The Chem participants learned four classes of organic chemistry stimuli (see Figure 1) adapted from two Chemistry textbooks (Baker, Rizzo, & Engel, 2010; Rittner & Bailey, 2005). The set of stimuli included names of four chemical structures (A stimuli), diagrams of these four structures (B stimuli), and descriptions of these four structures (C stimuli). Each grouping of name, picture, and function served as a class. The Dev Psych
participants learned four classes of developmental psychology research design stimuli adapted from Morris and Maisto (2005; see Figure 2). The set of stimuli included names of four research designs (A stimuli), diagrams of the four designs (B stimuli), and descriptions of the four designs (C stimuli). Each grouping of name, picture, and function served as a class. The stimuli used Times New Roman font, ranging from 16 pt to 33 pt font size.

All programs were written using Visual Basic. The computer tutorial administered all training and testing blocks (units) and recorded all participant responses. During all training and testing blocks in the EBI tutorial, one sample stimulus appeared in the center of the top portion of the screen and remained on throughout the trial. Once a participant clicked the sample stimulus (an observing response), four comparison stimuli (white or experimental stimuli) appeared in a row at the bottom of the screen. The computer tutorial randomized the presentation of trials and placement of comparison stimuli. A counter on the right-hand side of the screen provided visual feedback, which during training blocks displayed the mastery criterion (e.g., 12 consecutive correct responses) and the number of consecutive trials completed correctly. During testing blocks, the counter displayed the total number of test trials and how many trials the participant had already completed. The tutorial provided auditory accuracy feedback during all training trials, but not during testing trials. Following a correct response, the computer played an ascending tone (“chime” in Windows 10©) and the counter increased by one. Following an incorrect response, the computer played a descending tone (“chord” in Windows 10©) and the counter reset to 0.

A paper-based demographics questionnaire was administered prior to the tutorial (see Appendix B). The questionnaire assessed participants’ prior experiences with psychology and chemistry courses, experience in a research lab, and interest in pursuing a graduate degree.
Participants indicated responses to questions on the demographic questionnaire by circling either “Yes” or “No.” Questions 1 and 2 helped assess whether a participant’s data were to be excluded from analyses. Participants who reported taking the course relevant to their tutorial (i.e., General Psychology for the DevPsych condition and organic chemistry for the Chem condition) but scored at least 35.42% (17/48) on their tutorial’s 3Mix pretest (described below) were excluded from analyses. Data from participants who scored at least 35.42% on their tutorial’s 3Mix pretest but did not report having taken the course relevant to their tutorial were retained.

A paper-based exit survey (see Table 3) was administered after the completion of the tutorial. The exit survey assessed participant ratings of: 1) the importance of learning the concepts presented in this study and 2) motivation to learn the content presented in the respective tutorial.

**Design.** This study used a between-subjects pretest-posttest control group design (Campbell & Stanley, 1963). The experimenter assigned participants to conditions using block randomization (Urbonaik & Plous, 2014) at the start of the experiment, with two participants per block, assigned to either the Chem or Dev Psych conditions. The Chem group learned organic chemistry stimulus classes. Participants in the DevPsych condition learned developmental psychology research design stimulus classes.

**Procedure.** When the participant entered the experimental setting, the researcher obtained informed consent from the participant. The researcher explained all risks and benefits of participation. She explained that the participant had up to two hours to complete the procedure, but correct responding and quicker completion of the computer-based instruction would reduce time spent in the experimental setting. She also explained that there was a possibility that the
participant would be learning material that would be covered in the General Psychology course (see Appendix C).

Figure 3 depicts the procedure used in Experiment 1. After a participant provided consent, the experimenter administered a demographics questionnaire. The participant responded to each question by circling either “Yes” or “No.” After the participant completed the demographics questionnaire, the researcher started the appropriate computer program for the participant. First, the computer program administered a brief tutorial to orient the learner to its features. The tutorial then explained response requirements (click on the correct comparison given a sample) and feedback. Last, the tutorial administered two training blocks and a test of equivalence class formation regarding two classes that involved a picture, an Italian name, and a Chinese name for both orange and apple. The tutorial was identical to the one described by Fienup et al. (2015).

The next part of each computer program was the first administration of the 3Mix test. The 3Mix test was composed of 48 trials that assessed each baseline (A→B, A→C) and emergent relation (B→A, C→A, B→C, C→B) from all four classes within a stimulus set twice in a randomized order. The 3Mix pretest assessed baseline class-consistent responding (i.e., the degree to which participants matched stimuli from the same class together in the absence of training). No accuracy feedback was delivered during the 3Mix pretest. Following the completion of the 3Mix pretest, the computer program administered training and testing according to the simultaneous protocol (SIM; Fields, Reeve, Adams, Brown, & Verhave, 1997; see Figure 3). The SIM protocol involves mastering all training relations consecutively (A→B, A→C), followed by a single test phase (involving testing of all baseline and relations that are expected to emerge). A completion of the following sequence of events was considered one
cycle: A→B training, A→C training, and 3Mix posttest. On every training and testing trial, a participant had to select the correct comparison stimulus when presented with the sample stimulus. The tutorial trained conditional discriminations using a fading procedure, which has been shown to reduce the percentage of training errors committed (Brodsky & Fienup, in preparation). The fading procedure introduced each novel discrimination one at a time, with the S+ at full intensity at the onset of the trial and the text of the S-s fading in from white to grey to black. Across fading steps, three S-s were presented, but the intensity of the text relative to the background increased.

Each training phase was further subdivided into training stages (see Table 1), such that simple discriminations were introduced one at a time to reduce potential training errors. The A→B training phase was subdivided into the following seven sequential training stages: Class 1 (A1→B1); Class 2 (A2→B2); Classes 1 and 2 (alternating randomized presentations of A1→B1 and A2→B2 trials); Class 3 (A3→B3); Classes 1, 2, and 3 (alternating randomized presentations of A1→B1, A2→B2, and A3→B3 trials); Class 4 (A4→B4); and Classes 1, 2, 3, and 4 (alternating randomized presentations of A1→B1, A2→B2, A3→B3, and A4→B4 trials). Examples of the trial types presented in stages 3, 5, and 7 are presented in Figures 5, 6, 8, and 9, respectively.

Stages 1, 2, 4, and 6, which each taught only one relation (e.g., A1→B1), were further subdivided into three steps each (see Figure 4 and first row of Table 1). Over the course of these three steps, the incorrect comparisons faded in, and the mastery criterion for each of these steps was 2 consecutive correct trials. During the first step, the S+ was presented at full intensity, and the three S-s were presented as white stimuli (e.g., Fading 1-1). During the second step, the S+ was again presented at full intensity, while the three S-s were presented with grey text (e.g.,
During the third step, the S+ and S-s were all presented at their full intensity (e.g., Fading 1-3). Stages that taught more than one relation at a time were not further subdivided into steps, but had a mastery criterion of 12 consecutive correct responses. The seven stages were then repeated for A→C training. Following mastery of all A→C relations, the computer program administered the 3Mix posttest to each participant. If the participant mastered the 3Mix posttest (i.e., scored at least 90%), the tutorial terminated. If the participant failed the 3Mix posttest (i.e., scored below 90%), the tutorial reset to A→B training.

Following the completion of the final 3Mix posttest, the computer program presented a screen that gave the participant posttest feedback. If the participant scored under 90%, the computer tutorial presented a screen that read, “You did not score 90%. After completing the training you will have another opportunity to take the test.” If the participant scored at least 90% on the 3Mix posttest, the computer tutorial presented a screen that read, “You have completed the computer tutorial! Please leave this screen up and go get the research assistant.” Each participant who passed the 3Mix posttest was then asked to complete a paper-based exit survey, which asked the questions listed in Table 3. After completing the exit survey, the participant was dismissed from the experimental setting.

**Results and Discussion**

**Dependent Measures.** The following measures were calculated based on a participant’s first training and testing cycle: 3Mix pretest score, 3Mix posttest score, yield, number of training trials, percentage of training errors, and cycle time. The 3Mix pretest score assessed baseline responding to the Chem and DevPsych procedures. The following measures assessed the effectiveness of the Chem and DevPsych procedures: 3Mix posttest score and yield. The following measures assessed the efficiency of the Chem and DevPsych procedures: number of
training trials, percentage of training errors, and cycle time. Analyses were limited to data collected from the first training cycle to remove the influence of continued training on the dependent measures of interest. Exit survey data assessed the social validity of the Chem and DevPsych procedures.

Baseline responding was assessed by comparing differences in responding produced by the Chem and DevPsych conditions on the administration of the 3Mix pretest. Scores on the 3Mix pretest were calculated as follows: The researcher summed the number of correct responses emitted during the test, divided the sum by the total number of questions on the test (48), and multiplied this value by 100 to obtain a percentage. This calculation was repeated for each participant.

Following training, the 3Mix posttest score assessed how effectively participants formed equivalence classes after one cycle of training, and testing yield assessed the percentage of participants who formed equivalence classes on the first administration of the 3Mix posttest. Yield on the 3Mix test was calculated by summing the total number of participants in one condition who scored at least 43 out of 48 (90%) on the first administration of the 3Mix test, and then dividing this value by the total number of participants in a particular condition.

Number of training trials was calculated by summing the number of training trials required for a participant to reach mastery criterion in the A→B and A→C training phases. Percentage of training errors was calculated by summing the total number of correct trials engaged in by a participant during the first administration of the training, dividing this value by the sum of all trials the participant completed during the first administration of training, multiplying this value by 100 to obtain a percentage of correct responses, and then subtracting the percentage from 100 to obtain a percentage of errors. Cycle time was calculated by summing
the amount of time a participant required to complete the following phases: A→B training, A→C training, and the 3Mix posttest.

Social validity was calculated using data collected from the exit survey (see Table 3). Responses to questions 1 and 2 were calculated irrespective of condition. Responses to question 1 were calculated by summing the score each participant assigned to question 1 and dividing this value by the total number of participants who completed the question. This process was repeated for question 2. Responses to questions 3 and 4 were calculated per condition. Responses to questions 3 and 4 were each calculated by adding the scores a participant assigned to the particular question in each condition and then dividing each sum by the number of participants in that condition.

Effect sizes for the independent samples t-tests were calculated using Hedges’ g. Hedges’ g was selected to calculate effect sizes because this calculation corrects for unequal sample sizes. Values below 0.50 were considered small, values between 0.5 and 0.8 were considered medium, and values above 0.8 were considered large (Field, 2009).

Means, standard deviations, and sample sizes for 3Mix pretest scores, 3Mix posttest scores, number of training trials, percentage of training errors, and cycle time are listed in Table 2. Means, standard deviations, and distributions for exit survey data are presented in Table 3.

**Pretest Responding**

Differences in 3Mix pretest scores were evaluated using an independent samples t-test. Data for the 3Mix pretest measure are presented in Figure 9. There was no effect of condition on 3Mix pretest score, \( t(23.50)=-1.07, p=.32 \). Participants demonstrated baseline responding \((M=31.42\%\text{ for Chem, }M=35.83\%\text{ for DevPsych})\) that was only slightly higher than chance-level responding \((25\%)\).
**Effectiveness**

Effectiveness (i.e., whether participants formed equivalence classes) was evaluated based on two measures: Average 3Mix posttest score and posttest yield. Data for the 3Mix posttest measure are presented in Figure 10. Chem participants scored significantly higher on the 3Mix posttest than did DevPsych participants, \( t(14.56)=2.23, p<.05 \) (medium effect, \( g=0.77 \)).

Differences in yield were evaluated using a Chi-Square test of Independence. As seen in Figure 11, a significantly higher percentage of participants passed the final test of equivalence in the Chem condition (100.00%) than in the DevPsych condition (66.67%), \( \chi^2=8.73, p<.05 \).

**Efficiency**

Efficiency was evaluated based on number of training trials, percentage of training errors, and cycle time. Training trial data are presented in Figure 12. DevPsych participants required significantly more training trials to complete the first administration of training than did Chem participants, \( t(17.82)=-2.31, p<.05 \) (large effect, \( g=0.82 \)).

Training error data are also presented in Figure 12. DevPsych participants committed a significantly higher percentage of training errors than did Chem participants, \( t(15.62)=-2.89, p<.05 \) (large effect, \( g=1.00 \)).

Cycle time data are presented in Figure 13. Chem participants required significantly less time to complete a cycle than did DevPsych participants, \( t(25)=4.17, p<.05 \) (large effect, \( g=1.61 \)).

**Social Validity**

Descriptive statistics for social validity data are listed in Table 3. A within-subjects analysis was conducted using a Wilcoxon Signed Ranks test to determine whether there was a difference in how important participants thought it was to learn developmental psychology
(question 1) or chemistry (question 2) for their General Psychology course. Participants thought it was more important to learn developmental psychology for their General Psychology course than it was to learn chemistry for their General Psychology course, \( Z = -4.37, p < .05 \).

A Mann-Whitney U test was conducted to determine whether there were differences in responding between conditions on the responses to exit survey question 3 that asked participants if it was important for General Psychology students to learn the concepts presented in their tutorial and to the exit survey question that asked participants if they wanted to learn the concepts presented in their tutorial. There was no difference between the two conditions on how important participants thought it was to learn the concepts presented in their tutorial, \( U = 67.00, p = .28 \). There was a significant difference between the two conditions on how much participants said they wanted to learn the concepts presented in their tutorial, \( U = 46.00, p < .05 \). Participants in the DevPsych condition wanted to learn the concepts presented in their tutorial more than the participants in the Chem condition did.

**Summary**

There were no pretest differences between the Chem and DevPsych conditions. However, the Chem tutorial produced more effective and efficient outcomes. Participants agreed that it was more important to learn DevPsych concepts than Chem concepts for a General Psychology course. DevPsych participants said they wanted to learn the concepts presented in their tutorial more than did Chem participants, but there was no difference in how important participants in the two conditions thought it was to learn the particular concepts presented in their tutorial. Therefore, participants in the current study may have been aware that it is more important to learn one of these topics over the other, but exit survey data demonstrated that the relevance of course-related content may not have been apparent to the DevPsych participants.
Experiment 2

Experiment 1 determined that participants formed Chem equivalence classes more efficiently and effectively than they formed DevPsych classes. The Chem stimuli required fewer trials to complete training, committed 33% fewer training errors than did DevPsych participants, took less time to complete training and testing, and produced larger 3Mix posttest scores. Experiment 2 tested the effects of stimulus meaningfulness and educational relevance on equivalence class formation. The SIM protocol was used rather than the STC protocol as previous research shows that the STC protocol can increase yield to 100% (Fienup et al., 2015), which may obfuscate differences between conditions (Fields et al., 1997) thus reducing the sensitivity of measures. The Chem stimuli were used for the meaningful/ non-educationally relevant condition. The DevPsych stimuli were used for the meaningful/ educationally relevant condition. A new condition, Arbitrary, was introduced to determine whether meaningful stimuli produce more efficient and effective outcome than do arbitrary stimuli. The Arbitrary condition presented abstract stimuli drawn from Steele and Hayes (1991). Additionally, the current study included motivating operation procedures in an effort to increase the reinforcing value of learning DevPsych content (an EO) and decrease the reinforcing value of learning Chem and Arbitrary content (an AO).

If Experiment 2 found that DevPsych participants scored higher on efficiency and effectiveness measures, outcomes from Experiment 1 would help rule out stimulus difficulty alone as the source of differences.

Method

Participants and Setting. Fifty-one General Psychology students participated in this study in the same setting as in Experiment 1. They were recruited from the same pool used in
Experiment 1. The academic demographics data collected for the pool were as follows: 33% of participants planned to major in Psychology, 4% of participants planned to major in Chemistry, 20% had had experience in a research lab, 20% were interested in pursuing a research career, 10% were interested in getting a graduate degree in Chemistry or a related field, and 37% were interested in pursuing a career in the health professions.

**Stimuli, Software, and Materials.** The materials used in Experiment 2 were identical to those used in Experiment 1, with the exception of the addition of the arbitrary stimulus set. The stimuli for the Arbitrary condition, shown in Figure 14, were obtained from Steele and Hayes (1991).

**Design.** Experiment 2 used a three-group pretest-posttest control group design.

Participants were assigned to conditions using the same procedure as in Experiment 1.

**Procedure.** The procedure was the same as in Experiment 1 with a few exceptions (see Figure 15). First, the conditional discrimination training used in Experiment 2 was component conditional discrimination training (identical to 3SIM condition reported in Experiment 1 of Fienup et al., 2015), in which relations were trained in three stages (see Table 1). Stage 1 trained the first two relations (A1 → B1 and A2 → B2 in A → B training; A1 → C1 and A2 → C2 in A → C training) together by randomizing and presenting these two trial types until a participant responded correctly on 12 consecutive trials. Figure 5 depicts the two trial types presented in Stage 1. Stage 2 trained the next two relations (A3 → B3 and A4 → B4 in A → B training; A3 → C3 and A4 → C4 in A → C training) together by randomizing and presenting these two trial types until a participant responded correctly on 12 consecutive trials. Figure 7 depicts the two trial types presented in Stage 2. Stage 3 trained all four relations (A1 → B1, A2 → B2, A3 → B3, and A4 → B4 in A → B training followed by A1 → C1, A2 → C2, A3 → C3, and A4 → C4 in A → C training).
training) together by randomizing and presenting these four trial types until participant responded correctly on 12 consecutive trials. Figure 8 depicts the four trial types presented in Stage 3.

Second, the researcher included a MO manipulation, administered prior to the 3Mix pretest, in an effort to make salient the educational relevance of the concepts presented in each tutorial (see Figure 15). At the beginning of the tutorial, a participant was presented with statements about the concepts s/he was to learn. Each statement was directly followed by a question. The participants had to respond correctly to two consecutive presentations of each question to advance in the tutorial (described in detail below). If a participant responded incorrectly to either trial, a correction procedure was used in which the corresponding statement and question were re-presented. The rules presented were acquired with relatively few errors.

*DevPsych MO manipulation.* The DevPsych MO manipulation was intended to increase the momentary potency of the consequences presented in the DevPsych tutorial. Each DevPsych participant was told, “In this tutorial you will learn about developmental psychology research designs. This information will be covered in your Psych 101 course.” The participant then had to reply to the following question: “Will this tutorial cover material taught in your Psych 101 course?” The participant was presented with the following answer choices: “Yes” and “No.” The participant was required to reply by clicking “Yes,” for both presentations of the question. Next, the participant was presented with the following statement: “In this tutorial, you will be learning educational concepts.” The participant then had to reply to the following question: “Will you be learning educational or non-educational information in this tutorial?” The participant was presented with the following answer choices: “Educational” and “Non-educational.” The participant was required to reply by clicking “Educational,” for both presentations of the
question. Finally, the participant was shown the following statement: “In this tutorial, you will be learning academic concepts that will be taught in your Psych101 course.” The participant was then asked, “What type of concepts will you be learning in this tutorial?” The participant was presented with the following answer choices: “Non-educational concepts not covered in your Psych 101 course,” “Educational concepts not covered in your Psych 101 course,” and “Educational concepts that are covered in your Psych 101 course.” The participant was required to reply by clicking, “Educational concepts that are covered in your Psych 101 course,” for both presentations of the question.

***Chem MO Manipulation.*** The Chem MO manipulation was intended to reduce the momentary potency of the consequences presented in the Chem tutorial. Each Chem participant was told, “In this tutorial you will learn about chemical structures. This information will not be covered in your Psych 101 course.” The participant then had to reply to the following question: “Will this tutorial cover material taught in your Psych 101 course?” The participant was presented with the following answer choices: “Yes” and “No.” The participant was required to reply by clicking “No,” for both presentations of the question. Next, the participant was presented with the following statement: “In this tutorial, you will be learning educational concepts.” The participant then had to reply to the following question: “Will you be learning educational or non-educational information in this tutorial?” The participant was presented with the following answer choices: “Educational” and “Non-educational.” The participant was required to reply by clicking “Educational,” for both presentations of the question. Finally, the participant was shown the following statement: “In this tutorial, you will be learning academic concepts that will not be taught in your Psych101 course.” The participant was then asked, “What type of concepts will you be learning in this tutorial?” The participant was presented with
the following answer choices: “Non-educational concepts not covered in your Psych 101 course,” “Educational concepts not covered in your Psych 101 course,” and “Educational concepts that are covered in your Psych 101 course.” The participant was required to reply by clicking, “Educational concepts not covered in your Psych 101 course,” for both presentations of the question.

**Arbitrary MO Manipulation.** The Arbitrary MO manipulation was intended to decrease the momentary potency of the consequences presented in the Arbitrary tutorial. Each Arbitrary participant was told, “In this tutorial you will learn non-educational concepts. This information will not be covered in your Psych 101 course.” The participant then had to reply to the following question: “Will this tutorial cover material taught in your Psych 101 course?” The participant was presented with the following answer choices: “Yes” and “No.” The participant was required to reply by clicking “No,” for both presentations of the question. Next, the participant was presented with the following statement: “In this tutorial, you will be learning non-educational concepts.” The participant then had to reply to the following question: “Will you be learning educational or non-educational information in this tutorial?” The participant was presented with the following answer choices: “Educational” and “Non-educational.” The participant was required to reply by clicking “Non-educational,” for both presentations of the question. Finally, the participant was shown the following statement: “In this tutorial, you will be learning non-educational concepts that will not be taught in your Psych 101 course.” The participant was then asked, “What type of concepts will you be learning in this tutorial?” The participant was presented with the following answer choices: “Non-educational concepts not covered in your Psych 101 course,” “Educational concepts not covered in your Psych 101 course,” and “Educational concepts that are covered in your Psych 101 course.” The participant was required
to reply by clicking, “Non-educational concepts **not** covered in your Psych 101 course,” for both presentations of the question.

**Results and Discussion**

**Dependent Measures.** The dependent measures used in Experiment 2 were the same as those used in Experiment 1. Additionally, Eta-squared was used to calculate effect sizes for the one-way ANOVA comparing 3Mix pretest outcomes. Interpretations for Eta eta-squared were as follows: Values below .09 were considered small, values between .09 and .25 were considered medium, and values above .25 were considered large (Field, 2009).

Means, standard deviations, and sample sizes for 3Mix pretest scores, 3Mix posttest scores, number of training trials, percentage of training errors, and cycle time are listed in Table 4. Means, standard deviations, and distributions for exit survey data are presented in Table 5.

**Pretest Responding.** Data for the 3Mix pretest measure are presented in Figure 16. Group differences in 3Mix pretest score were evaluated using a one-way ANOVA with post-hoc testing using the Bonferroni method. There was a significant effect of condition on 3Mix pretest, $F(2,50)=4.34, p<.05$ (medium effect, $\eta^2=0.15$). There was no difference on the 3Mix pretest between the Arbitrary and Chem conditions, $p=1.00$. However, Arbitrary participants scored significantly lower on the 3Mix pretest than did DevPsych participants, $p<.05$. There was no difference between Chem and DevPsych 3Mix pretest scores, $p=0.60$. However, participants in all three conditions demonstrated low baseline performance ($M=28.87\%$ for Arbitrary, $M=31.15\%$ for Chem, and $M=41.54\%$ for DevPsych).

**Effect of Meaningfulness.** To analyze the effects of meaningfulness, outcomes of the Arbitrary condition (all non-meaningful stimuli) were compared to the aggregated outcomes of the Chem and Dev Psych condition (both consisted of meaningful stimuli). A priori contrasts
were used to evaluate 3Mix posttest scores, training trials, percentage of training errors, and cycle time. Means, standard deviations, and sample sizes are listed in Table 4.

**Effectiveness.** Effectiveness was evaluated by examining 3Mix posttest scores, presented in Figure 17. Participants in the Arbitrary condition obtained significantly lower 3Mix posttest scores than did participants in the meaningful conditions, \( t(22.02)=14.80, p<.05 \) (large effect, \( g=4.35 \)).

**Efficiency.** Training trial data for the Arbitrary and meaningful conditions are presented in Figure 18. Participants in the Arbitrary condition required significantly more trials to complete training than did participants in the meaningful conditions, \( t(48)=-2.33, p<.05 \) (medium effect, \( g=0.74 \)).

Training error data are presented in Figure 18. Participants in the Arbitrary condition committed a significantly greater percentage of training errors than did participants in the meaningful conditions, \( t(48)=-2.13, p<.05 \) (medium effect, \( g=0.67 \)).

Cycle time data are presented in Figure 19. Participants in the Arbitrary condition required significantly less time to complete the first cycle of training and testing than did participants in the two meaningful conditions, \( t(34.31)=3.74, p<.05 \) (large effect, \( g=0.82 \)).

**Effect of Educational Relevance.** To analyze the effects of educational relevance as a motivating operation, the outcomes of the Chem condition and Dev Psych condition were compared. Data from the Chem and Arbitrary conditions were not combined, despite both conditions containing non-educationally relevant stimuli, because this analysis was purely aimed at determining the effect of educational relevance, without confounding the data with the effect of arbitrariness. Differences in 3Mix posttest score, training trials, percentage of training errors,
and cycle time were evaluated by using a priori contrasts. Means, standard deviations, and sample sizes are listed in Table 4.

**Effectiveness.** As shown in Figure 17, participants in the DevPsych condition produced significantly higher 3Mix posttest scores than did participants in the Chem condition, $t(23.31)=2.26, p<.05$ (medium effect, $g=0.70$). This result contrasts with Chem participants scoring higher on the 3Mix posttest than DevPsych participants in Experiment 1.

Differences in the yield data for the Chem and DevPsych conditions (shown in Figure 20) were evaluated using a Chi-Square test of Independence. A greater percentage of participants passed the final test of equivalence in the DevPsych (94.12%) condition than in the Chem condition (75.00%), $\chi^2=4.34, p<.05$. This effect contrasts with yield outcomes of Experiment 1, in which participants in the Chem condition (100.00%) were more likely to pass the final test of equivalence than were participants in the DevPsych condition (66.67%). Thus, in Experiment 2 as compared to Experiment 1, participants showed an increase in yield for the DevPsych condition and a decrease in yield for the Chem condition.

**Efficiency.** Training trial data for the Chem and DevPsych conditions are presented in Figure 19. In contrast with Experiment 1, there was no difference in the number of trials that participants in the Chem and DevPsych conditions required to complete training, $t(48)=1.60, p=.12$.

Also as shown in Figure 18, in contrast to Experiment 1, there was no difference between participants in the DevPsych and Chem condition in percentage of training errors they committed, $t(48)=1.07, p=.29$. 
As shown in Figure 19, participants in the Chem condition required significantly less time to complete the first cycle of training and testing than did participants in the DevPsych condition, \( t(25.39) = 3.52, p < .05 \) (large effect, \( g = 1.21 \)).

**Social validity.** Social validity data are displayed in Table 5. A within-subjects analysis was conducted using a Wilcoxon Signed Ranks test to determine whether there was a difference in how important DevPsych and Chem participants thought it was to learn developmental psychology (question 1 of the exit survey) or chemistry (question 2 of the exit survey) for their General Psychology course. Participants thought it was more important to learn developmental psychology for their General Psychology course than it was to learn chemistry for their General Psychology course, \( Z = -4.73, p < .05 \).

A Mann-Whitney U test was conducted to determine whether there were differences in responding between conditions on the responses to exit survey question 3 that asked participants if it was important for General Psychology students to learn the concepts presented in their tutorial and to exit survey question 4 that asked participants if they wanted to learn the concepts presented in their tutorial. DevPsych participants thought it was more important to learn the content presented in their tutorial than the importance that Chem participants ascribed to their tutorial, \( U = 76.00, p < .05 \). DevPsych participants also wanted to learn the concepts presented in their tutorial more than did the Chem participants, \( U = 58.50, p < .05 \).

**Summary**

The Arbitrary condition produced less effective performances than did the meaningful tutorials and required less time to complete training and testing than did participants in the meaningful conditions, but required more trials and committed a greater percentage of training errors than did participants in the meaningful conditions. Participants in the Arbitrary condition
completed trials at a faster rate, but with less accuracy than did participants in the meaningful conditions.

Within the meaningful conditions, DevPsych participants required more time to complete training than did the Chem participants, but they did so with the comparable number of training trials and percentage of training errors. Additionally, DevPsych participants scored higher on the first administration of the 3Mix posttest than did Chem participants, and social validity measures show that they were more motivated to learn the content presented in their tutorial.

**General Discussion**

Results from the current study show that arbitrary stimuli produce lower posttest scores relative to meaningful stimuli, and that educational relevance may function as a MO that affects equivalence class formation. In Experiment 1, in which there was no MO manipulation and the fading procedure was used during conditional discrimination training, participants who learned the educationally irrelevant equivalence classes did so more effectively (i.e., higher posttest scores and higher yield) and more efficiently (i.e., fewer number of training trials, lower percentage of training errors, shorter cycle time) than did participants who learned educationally relevant equivalence classes. The results of Experiment 1 demonstrate that stimulus difficulty cannot account for the higher posttest scores and yield that the educationally relevant stimuli produced in Experiment 2. Experiment 2 deliberately used a conditional discrimination training procedure shown to produce less efficient performance (Brodsky & Fienup, in preparation) and manipulated a MO across the Arbitrary, Chem, and DevPsych conditions. Experiment 2 demonstrated that participants learning meaningful stimuli did so more effectively (higher posttest scores) and efficiently (percentage of training errors) than did participants learning arbitrary stimuli. The MO caused participants learning educationally relevant stimuli to do so
more effectively (i.e., higher posttest scores and higher yield) than participants learning non-
educationally relevant stimuli.

Findings from this study contribute to the literature in a number of ways. The current study replicated the effect of meaningfulness on equivalence class formation demonstrated by Arntzen (2004), Fields et al. (2012), and Mensah and Arntzen (2016). Additionally, the current study extended the research on meaningfulness by comparing an entire set of arbitrary stimuli to two different sets of meaningful stimuli. In contrast, Arntzen (2004) and Fields et al. (2012) taught three 5-member classes which contained up to one meaningful stimulus per class, and Mensah and Arntzen (2016) taught three 5-member classes that contained between one and three classes with one meaningful stimulus per class. Furthermore, the meaningful stimuli used for comparison with arbitrary stimuli in previous research were limited to pictures of a church, crown, and mailbox. The current study expanded the comparison between arbitrary and meaningful stimuli to 48 novel meaningful stimuli that included names and definitions in addition to pictures of chemical structures and developmental psychology research methods. By comparing an entire set of arbitrary stimuli to two different meaningful stimulus sets and using novel meaningful stimuli, the current study demonstrates the generality of the effect of meaningfulness on equivalence class formation.

The current study also demonstrated that the effect of meaningfulness can be broken down further—there is a difference between the effect of meaningful/ non-educationally relevant stimuli and the effect of meaningful/ educationally relevant stimuli. This novel finding may have far-reaching implications on the manner in which equivalence research is conducted. The effect of educational relevance as a MO for equivalence class formation may provide an alternate explanation for why variables investigated in the basic and translational context may not have
produced similar results in the applied context, and why researchers may use procedural variations of EBI that produce low yields in basic research, such as the LS training structure (e.g., Fienup et al., 2016) and the SIM protocol (Pytte & Fienup, 2012), without necessarily producing low yields in the applied setting.

The impact of the current study’s findings are two-fold. First, an understanding of the effect of educational relevance may be beneficial in instructional settings. The findings here showed that the effects of educational relevance were large enough to counter the effects of stimulus difficulty. Such findings can benefit both students learning course-relevant content and classroom instructors designing EBI content for their students. If educational relevance, such as possibility of EBI stimuli appearing on an exam, serves as a MO for learning course-related content, then it may be possible to enhance classroom instruction through equivalence procedures even when students consider the content difficult to learn. The findings in the current study may also mean that instructors can use variants of EBI that have been shown to be less effective in basic research but are easier for researchers or instructors to program, because the effect of educational relevance may increase yields irrespective of the effects of other procedural variables. Second, findings from the current study call for a new line of equivalence research for college-level instructional applications—research investigating how all EBI training parameters interact with the effect of educational relevance to help participants form equivalence classes most effectively and efficiently. For example, basic investigations on nodal number (e.g., Fields et al., 1997), MTS delay (Arntzen, 2006), training structure (Arntzen & Holth, 1997), and translational investigations on training protocol (e.g., Fienup et al., 2016) should be repeated with college-level students using course-relevant stimuli to see which procedural variants, when
used with meaningful/ educationally relevant stimuli, reduce instructional time and produce the highest yields.

One limitation of the current study was that the demographics questionnaire and exit survey were not validated with psychometric analyses prior to use in Experiments 1 and 2. Validating the questionnaires would help identify whether the questions used in the current study were capturing the constructs they were intended to measure. For example, in Experiment 1, 0% of participants planned on majoring in Chemistry but 11% planned on obtaining a graduate degree in Chemistry or a related field. Both questionnaire items (questions 4 and 7 on the demographics questionnaire, see Appendix B) were intended to assess a participant’s interest in chemistry, but there were discrepancies in responding on these two questions. Additionally, demographics questionnaire items may not have been sensitive to participant learning histories with psychology and chemistry. It is possible that participants in the current study were briefly previously exposed to organic chemistry content in their high school course, but it is less likely they have come into contact with developmental psychology research methods in high school (unless they were enrolled in an Advanced Placement Psychology course). Future studies should develop and validate questionnaires that better assess motivation for learning EBI content. A second limitation of the current study was that the current study used a between-subjects design. A within-subjects design would have been more powerful, allowing for the demonstration of an effect over and above any individual differences. However, such an experimental design is not advisable in equivalence research. Due to learning set (Catania, 2007), completing one EBI tutorial enhances learning outcomes such as reduced training time (Fienup et al., 2015) and higher posttest yields (Buffington, Fields, & Adams, 1997) on subsequent tutorials. Future studies should examine how educational relevance affects concept formation when concepts vary
across a physical dimension using a within-subject design. Such a procedure would involve concept formation using discrimination training (Herrnstein, Loveland, & Cable, 1976; Siegel & Honig, 1970).

Furthermore, it should be noted that the procedures used in the current study were carefully chosen to implement the independent variables discussed given available resources. The Arbitrary stimuli were chosen because they had been used in previous research (Steele & Hayes, 1991) and during pilot testing in preparation for the current study. The particular chemistry stimuli used in this experiment were chosen to minimize differences in picture and function that could have led participants to rely on components within a stimulus, rather than the stimulus itself (i.e., stimulus over-selectivity), to form equivalence classes. Developmental psychology stimuli were chosen for the educationally relevant group of the current study due to the timeline of the General Psychology course from which participants were drawn—that is, the experimenter aimed to run the current study before students were exposed to the related instruction in class, but with enough time to recruit a large enough sample that would allow for the demonstration of effects. This participant pool eliminated a number of topics that could also have been used in the educationally relevant condition, such as Learning, Neuroscience, and general Research Methods in psychology. Future research should be conducted to determine whether using stimulus sets differing along dimensions such as number of sentences, reinforcement history with the words used in the name and definition, and number of syllables in the stimuli affects outcomes such as yield or percentage of training errors. Additionally, due to Visual Basic program constraints, the tutorial developed for this study does not allow for the remote extraction of data. Therefore, tutorials had to be completed in the laboratory setting rather
than online using distance education software such as Blackboard, Moodle, or Google Classrooms.

Future research should administer EBI tutorials using distance education to determine the generality of the current study’s findings when the intervention is implemented in the context of a course. The implementation of the independent variables used in the current study could have also varied on a number of other parameters, such as: mastery criterion for training and testing, training structure, training protocol, and participant pools. Future research should use different combinations of other EBI parameters in conjunction with educationally relevant stimuli to determine the generality of educational relevance effect in EBI research. Further, this experiment should be replicated with students enrolled in an organic chemistry course to determine whether the effect of educational relevance demonstrated in the current study will still hold when organic chemical stimuli are educationally relevant rather than developmental psychology stimuli. The effects of educational relevance and meaningfulness should be carefully studied under conditions different from those used in the current study to determine the generality of the findings presented here. Future research should also carefully examine the parameters of a MO manipulation to determine how to maximize educational gains for students completing EBI tutorials. In the current study, the MO manipulation included a sequence of three statements about the educational relevance and meaningfulness of the stimuli used in the tutorial, each followed by a question reiterating the information in the statement. Researchers should identify whether a MO manipulation involving more questions would have a stronger effect on equivalence class formation, or whether one statement/question pair alone would have been enough to produce the effects observed in the current study. Furthermore, would statements alone be sufficient to affect behavior, or was the question also a necessary component? Are there
variations of the wording used that would have created more or less effective and efficient outcome? Additionally, it is important to determine how using such a manipulation would function when a student completes a sequence of EBI tutorials for a course. Does the effect of the manipulation maintain over time, or does the wording of the manipulation need to change with the tutorial in order to avoid habituation to the stimuli?

Future directions should also include determining which aspect of the applied context affects the application of basic findings to educational uses of EBI. Are there greater outcome disparities along the continuum of educational relevance if students complete tutorials just before an exam on the content, rather than a few months before? Additionally, it should be noted that learning educationally relevant stimuli in Experiment 2 did not produce 100% yield. Therefore future research should identify EBI training components that increase yields to even higher levels. Changing the context of college-level equivalence research to consider educational relevance will help researchers determine which components of instruction affect equivalence class-consistent responding enough to warrant implementation, and which components are less important to program for when adapting EBI for classroom use. Determining which variables impact equivalence class formation in the applied context may streamline instructional design of EBI, in turn facilitating its adoption in college classrooms and producing better-trained graduates.
Table 1.

**Stages and Steps of Conditional Discrimination Training**

<table>
<thead>
<tr>
<th>Classes Trained</th>
<th>Relations Trained</th>
<th>Mastery Criterion</th>
<th>Fading Stage-Step</th>
<th>Component Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A1→B1 (white S-s)</td>
<td>2</td>
<td>1-1</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>A1→B1 (grey S-s)</td>
<td>2</td>
<td>1-2</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>A1→B1 (black S-s)</td>
<td>2</td>
<td>1-3</td>
<td>n/a</td>
</tr>
<tr>
<td>2</td>
<td>A2→B2 (white S-s)</td>
<td>2</td>
<td>2-1</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>A2→B2 (grey S-s)</td>
<td>2</td>
<td>2-2</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>A2→B2 (black S-s)</td>
<td>2</td>
<td>2-3</td>
<td>n/a</td>
</tr>
<tr>
<td>1 and 2</td>
<td>A1→B1, A2→B2</td>
<td>12</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>A3→B3 (white S-s)</td>
<td>2</td>
<td>4-1</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>A3→B3 (grey S-s)</td>
<td>2</td>
<td>4-2</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>A3→B3 (black S-s)</td>
<td>2</td>
<td>4-3</td>
<td>n/a</td>
</tr>
<tr>
<td>1, 2, and 3</td>
<td>A1→B1, A2→B2, A3→B3</td>
<td>12</td>
<td>5</td>
<td>n/a</td>
</tr>
<tr>
<td>4</td>
<td>A4→B4 (white S-s)</td>
<td>2</td>
<td>6-1</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>A4→B4 (grey S-s)</td>
<td>2</td>
<td>6-2</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>A4→B4 (black S-s)</td>
<td>2</td>
<td>6-3</td>
<td>n/a</td>
</tr>
<tr>
<td>3 and 4</td>
<td>A3→B3, A4→B4</td>
<td>12</td>
<td>n/a</td>
<td>2</td>
</tr>
<tr>
<td>1, 2, 3, and 4</td>
<td>A1→B1, A2→B2, A3→B3, A4→B4</td>
<td>12</td>
<td>7</td>
<td>3</td>
</tr>
</tbody>
</table>

*Note.* The table displays stages and steps for A→B conditional discrimination training, which are repeated for A→C training. The numbers under the Fading and Comp columns correspond to the order in which training stages/steps were administered. The numbers under the mastery criterion column reflect the number of consecutive correct trials that a participant had to respond to correctly.
Table 2.

*Descriptive Statistics for Experiment 1 Tutorial Variables*

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Condition</th>
<th>N</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>3Mix Pretest Score</td>
<td>Chem</td>
<td>12</td>
<td>31.42%</td>
<td>8.02</td>
</tr>
<tr>
<td></td>
<td>DevPsych</td>
<td>15</td>
<td>35.83%</td>
<td>13.23</td>
</tr>
<tr>
<td>3Mix Posttest Score</td>
<td>Chem</td>
<td>12</td>
<td>98.61%</td>
<td>2.05</td>
</tr>
<tr>
<td></td>
<td>DevPsych</td>
<td>15</td>
<td>89.17%</td>
<td>16.24</td>
</tr>
<tr>
<td>Training Trials</td>
<td>Chem</td>
<td>12</td>
<td>128.55</td>
<td>9.68</td>
</tr>
<tr>
<td></td>
<td>DevPsych</td>
<td>15</td>
<td>146.87</td>
<td>28.65</td>
</tr>
<tr>
<td>Training Errors</td>
<td>Chem</td>
<td>12</td>
<td>1.41%</td>
<td>1.39</td>
</tr>
<tr>
<td></td>
<td>DevPsych</td>
<td>15</td>
<td>6.35%</td>
<td>6.43</td>
</tr>
<tr>
<td>Cycle Time</td>
<td>Chem</td>
<td>12</td>
<td>603.67 s</td>
<td>118.97</td>
</tr>
<tr>
<td></td>
<td>DevPsych</td>
<td>15</td>
<td>881.13 s</td>
<td>204.39</td>
</tr>
</tbody>
</table>
Table 3.

*Descriptive Statistics for Experiment 1 Exit Survey Responses*

<table>
<thead>
<tr>
<th></th>
<th>All (N=27)</th>
<th>Chem (n=12)</th>
<th>DevPsych (n=15)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Distribution</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>1. It is important for me to learn developmental psychology concepts for my General Psychology course.</td>
<td>0-0-2-9-16</td>
<td>4.52</td>
<td>0.64</td>
</tr>
<tr>
<td>2. It is important for me to learn chemistry for my General Psychology course.</td>
<td>5-5-12-5-0</td>
<td>2.63</td>
<td>1.01</td>
</tr>
<tr>
<td>3. It is important for General Psychology students to learn the concepts presented in this course.</td>
<td>2-3-7-0-0</td>
<td>2.42</td>
<td>0.79</td>
</tr>
<tr>
<td>4. I wanted to learn the concepts presented in this study.</td>
<td>1-1-5-4-1</td>
<td>3.25</td>
<td>1.06</td>
</tr>
</tbody>
</table>
Table 4.

*Descriptive Statistics for Experiment 2 Tutorial Variables*

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Condition</th>
<th>N</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>3Mix Pretest Score</td>
<td>Arbitrary</td>
<td>14</td>
<td>28.87%</td>
<td>10.96</td>
</tr>
<tr>
<td></td>
<td>Chem</td>
<td>20</td>
<td>31.15%</td>
<td>16.09</td>
</tr>
<tr>
<td></td>
<td>DevPsych</td>
<td>17</td>
<td>41.54%</td>
<td>10.66</td>
</tr>
<tr>
<td>3Mix Posttest Score</td>
<td>Arbitrary</td>
<td>14</td>
<td>47.17%</td>
<td>10.13</td>
</tr>
<tr>
<td></td>
<td>Chem</td>
<td>20</td>
<td>90.00%</td>
<td>13.43</td>
</tr>
<tr>
<td></td>
<td>DevPsych</td>
<td>17</td>
<td>97.18%</td>
<td>4.22</td>
</tr>
<tr>
<td>Training Trials</td>
<td>Arbitrary</td>
<td>14</td>
<td>123.64</td>
<td>28.65</td>
</tr>
<tr>
<td></td>
<td>Chem</td>
<td>20</td>
<td>96.70</td>
<td>21.16</td>
</tr>
<tr>
<td></td>
<td>DevPsych</td>
<td>17</td>
<td>110.82</td>
<td>31.92</td>
</tr>
<tr>
<td>Training Errors</td>
<td>Arbitrary</td>
<td>14</td>
<td>13.58%</td>
<td>6.43</td>
</tr>
<tr>
<td></td>
<td>Chem</td>
<td>20</td>
<td>8.55%</td>
<td>5.37</td>
</tr>
<tr>
<td></td>
<td>DevPsych</td>
<td>17</td>
<td>10.65%</td>
<td>6.19</td>
</tr>
<tr>
<td>Cycle Time</td>
<td>Arbitrary</td>
<td>14</td>
<td>600.86 s</td>
<td>147.60</td>
</tr>
<tr>
<td></td>
<td>Chem</td>
<td>20</td>
<td>671.10 s</td>
<td>170.93</td>
</tr>
<tr>
<td></td>
<td>DevPsych</td>
<td>17</td>
<td>947.41 s</td>
<td>283.19</td>
</tr>
</tbody>
</table>
## Descriptive Statistics for Experiment 2 Exit Survey Responses

<table>
<thead>
<tr>
<th></th>
<th>All (N=37)</th>
<th>Chem (n=20)</th>
<th>DevPsych (n=17)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Distribution</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>1. It is important for me to learn developmental psychology concepts for my General Psychology course.</td>
<td>1-1-5-10-20</td>
<td>4.27</td>
<td>0.99</td>
</tr>
<tr>
<td>2. It is important for me to learn chemistry for my General Psychology course.</td>
<td>4-13-10-8-2</td>
<td>2.75</td>
<td>1.09</td>
</tr>
<tr>
<td>3. It is important for General Psychology students to learn the concepts presented in this course.</td>
<td>5-4-4-3-4</td>
<td>2.85</td>
<td>1.50</td>
</tr>
<tr>
<td>4. I wanted to learn the concepts presented in this study.</td>
<td>1-3-9-3-4</td>
<td>3.30</td>
<td>1.13</td>
</tr>
<tr>
<td>Name (A)</td>
<td>Diagram (B)</td>
<td>Description (C)</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td>-----------------</td>
<td></td>
</tr>
<tr>
<td>Alcohol</td>
<td><img src="image" alt="Alcohol Diagram" /></td>
<td>A hydrocarbon derivative containing a hydroxyl group attached to a carbon atom.</td>
<td></td>
</tr>
<tr>
<td>Ketone</td>
<td><img src="image" alt="Ketone Diagram" /></td>
<td>An organic compound in which a carbonyl group is bound to two carbon atoms.</td>
<td></td>
</tr>
<tr>
<td>Carboxylic acid</td>
<td><img src="image" alt="Carboxylic Acid Diagram" /></td>
<td>Organic molecules with a carbonyl group in which the carbon is bonded to a hydroxyl group.</td>
<td></td>
</tr>
<tr>
<td>Amide</td>
<td><img src="image" alt="Amide Diagram" /></td>
<td>Derivative of ammonia in which one or more hydrogen atoms are replaced by the alkyl or aryl groups.</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 1. Stimuli used in the Chem condition.*
Figure 2. Stimuli used in the DevPsych condition.
Figure 3. Procedure used in Experiment 1. Mastery criteria for a phase, if applicable, are listed in parentheses. CC stands for “consecutive correct.” Relations trained in each phase are presented below the training/testing phase. Relations that are expected to emerge after training are marked with an asterisk (*). Baseline relations are marked with a superscripted B (B). Emergent relations are marked with a superscripted E (E). The sequence of relations listed was fixed during training phases. During testing phases, the relations shown were presented to the participant in a randomized fashion two times each.
Figure 4. Steps in the fading procedure. The top, middle, and bottom screenshots show the White, Grey, and Black steps, respectively (Stages 1, 2, 4, & 6 of Table 1 above).
Figure 5. The two trial types presented in Stage 3 of fading conditional discrimination training and Stage 1 of component conditional discrimination training, which trained relations A1→B1 and A2→B2.
Figure 6. The three trial types presented in Stage 5 of fading conditional discrimination training, which trained relations A1→B1, A2→B2, and A3→B3.
Figure 7. The two trial types presented in Stage 2 of component conditional discrimination training, which trained relations $A_3 \rightarrow B_3$ and $A_4 \rightarrow B_4$. 
Figure 8. The four trial types presented in Stage 7 of fading conditional discrimination training and Stage 3 of component conditional discrimination training, which trained relations A1→B1, A2→B2, A3→B3, and A4→B4.
Figure 9. Scores on the 3Mix pretest for the Chem and DevPsych conditions in Experiment 1.

Grey circles represent individual data points. Solid black lines represent condition means.
Figure 10. Scores on the 3Mix posttests for the Chem and DevPsych conditions in Experiment 1. Grey circles represent individual data points. Solid black lines represent condition means.
Figure 11. Percentage of participants reaching mastery criterion (90%) on the first administration of the 3Mix posttest in the Chem and DevPsych conditions in Experiment 1.
Figure 12. Number of training trials (top) and percentage of training errors (bottom) for the Chem and DevPsych conditions in Experiment 1. Grey circles represent individual data points. Solid black lines represent condition means.
Figure 13. Cycle time for the Chem and DevPsych conditions in Experiment 1. Grey circles represent individual data points. Solid black lines represent condition means.
Figure 14. Stimuli used in the Arbitrary condition drawn from Steele and Hayes (1991).
Figure 15. Procedure used in Experiment 2. Mastery criteria for a phase, if applicable, are listed in parentheses. CC stands for “consecutive correct.” Relations trained in each phase are presented below the training/testing phase. Relations that are expected to emerge after training are marked with an asterisk (*). Baseline relations are marked with a superscripted B ($^B$). Emergent relations are marked with a superscripted E ($^E$). The sequence of relations listed was fixed during training phases. During testing phases, the relations shown were presented to the participant in a randomized fashion two times each.
Figure 16. Scores on the 3Mix pretest for the Arbitrary, Chem, and DevPsych conditions in Experiment 2. Grey circles represent individual data points. Solid black lines represent condition means.
Figure 17. Scores on the 3Mix posttest for the Arbitrary, Chem, and DevPsych conditions in Experiment 2. Grey circles represent individual data points. Solid black lines represent condition means.
Figure 18. Number of training trials (top) and percentage of training errors (bottom) for the Arbitrary, Chem, and DevPsych conditions in Experiment 2. Grey circles represent individual data points. Solid black lines represent condition means.
Figure 19. Cycle time for the Arbitrary, Chem, and DevPsych conditions in Experiment 2. Grey circles represent individual data points. Solid black lines represent condition means.
Figure 20. Percentage of participants reaching mastery criterion (90%) on the first administration of the 3Mix posttest in the Chem and DevPsych conditions in Experiment 2.
Appendix A

Glossary of Terms

**LS (Linear Series):** Baseline relations are taught in sequence with multiple linking stimuli.

**MTO (Many-to-One):** A training structure in which many comparisons are trained to one sample stimulus (B → A and C → A).

**OTM (One-to-Many):** A training structure in one sample stimulus is trained to many comparisons (A → B and A → C).

**SIM (Serial Simultaneous):** A training protocol that teaches all baseline relations first, followed by one final test of equivalence.

**STC (Simple-to-Complex):** A training protocol that intermixes training and testing phases.

**Training protocol:** The manner in which training and testing phases are arranged.

**Training structure:** The arrangement of overlapping conditional discriminations.

**Yield:** The percentage of participants passing the final test of equivalence on their first attempt.
Appendix B
Demographics Questionnaire

Please circle one response for each of the questions listed below:

1. Have you previously taken a General Psychology course in high school or college?
   Yes  No

2. Have you previously taken an organic chemistry course?
   Yes  No

3. Are you planning on majoring in Psychology?
   Yes  No

4. Are you planning on majoring in Chemistry?
   Yes  No

5. Do you have experience working in a research laboratory?
   Yes  No

6. Are you interested in pursuing a research career?
   Yes  No

7. Are you interested in pursuing a graduate degree in Chemistry or a related field?
   Yes  No

8. Are you interested in pursuing a career in healthcare (E.g., doctor, nurse, physician’s assistant)?
   Yes  No
Appendix C

Script for Consent Procedure

- “You are being asked to participate in a research project conducted through Queens College CUNY. If you decide to participate, Queens College requires that you give your signed authorization to participate in this research project.”
- “We are piloting a computer program that teaches concepts. We are developing this tutorial to supplement college-level instruction in courses such as General Psych or Chemistry.”
- “You will be working on a computer program and you will be responding to question and answer situations.”
- “There are no known risks beyond what you would experience in a course here at Queens College. You may benefit by learning academic material that will be covered in your courses.”
- “There are alternatives to completing this research project such as completing other studies, attending colloquiums, or writing a research report. Here is a website you can go to that explains the alternatives” (point to website on p. 2 of consent form).
- “This study will take you between 1 and 2 hours to complete. You can leave when you are done with the study or after 2 hours have elapsed, whichever comes first.”
- “Do you have any questions?”
- “Read over the consent form at your leisure and sign at the end if you would like to participate. One copy of the consent form is for you to keep.”
References


discriminative, and a meaningless stimulus on equivalence class formation. *Journal of the
Experimental Analysis of Behavior, 97*, 163-181.

distance effects in equivalence classes. *Journal of the Experimental Analysis of Behavior, 64*,
129-145.

extension of equivalence classes from primary generalization gradients: The merger of
equivalence classes and perceptual classes. *Journal of the Experimental Analysis of
Behavior, 42*, 575-593.

Equivalence class formation: A method for teaching statistical interactions. *Journal of
Applied Behavior Analysis, 42*, 575-593.

using stimulus equivalence.*

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.

Effects of training protocols on equivalence class formation. *Journal of Applied Behavior
Analysis, 48*, 613-631.

*Journal of the Experimental Analysis of Behavior, 52*, 275-291.


