Effects of Equivalence-Based Instruction in Teaching Health Related Norwegian Concepts to Adults with a Foreign Background

Oana Pintilie, Torunn Lian, and Erik Arntzen
Oslo Metropolitan University

This study examined the effectiveness of Equivalence-Based Instruction (EBI) in teaching Norwegian health concepts to adults with a foreign background. In this stimulus control arrangement, we also wanted to see if using the participants' native languages would improve their performance. Eight adults were assigned to two groups, one group experienced text stimuli in Norwegian only and the other experienced some text stimuli in their respective, native languages. The participants underwent a training session where baseline conditional discriminations were established using a one-to-many training structure (AB, AC, and AD), followed by a test for emergent stimulus-stimulus relations. The A-stimuli were names of the four main body systems, the B-stimuli were pictures of the systems, and the C stimuli were definitions, descriptions or labels of the systems, and D stimuli were the names in different native languages. Overall, the results showed that seven out of eight participants formed equivalence classes. The results are in accordance with previous findings indicating that equivalence based technology seems to be a promising way to teach foreign language concepts. With regards to using native text stimuli, we did not succeed in differentiating the experimental condition. Limitations of the present experiment are discussed and directions for future research are suggested.

Keywords: stimulus equivalence, conditional-discrimination training, second language, adults, native language

---

The importance of mastering the language of the country you reside is emphasized within the political agenda established by the Norwegian government in the Introduction Law. Learning Norwegian is predicted to help immigrants in gaining access to education and work, as well as taking part in societal obligations or activities (Kunnskapsdepartementet, 2019). Not only does learning the primary, spoken language of the specific country give access to increased or better job opportunities, but it could also aid in society challenges, such as poverty (Kaida, 2013). The need for new methods in teaching adults with foreign background is increasingly brought into discussion as reports show that the current teaching format is neither adaptable nor effective for individuals with special needs, low or high level of education (Djuve et al., 2017). Therefore, the search for effective methods becomes even more relevant for the field of second language acquisition, as it could potentially meet the criteria for individualized instructional formats. Since Norwegian instruction for individuals with immigrant background is a priority and a requirement from the Norwegian government, we deemed this topic as valuable to test the effectiveness of

The present article is based on data collected as part of the first authors' master thesis. Correspondence concerning this article should be addressed to Oana Pintilie o.pintilie@gmail.com

stimulus control procedures in an applied setting within second language acquisition.

From a stimulus control perspective, second language acquisition requires words from the foreign language to become members of the same stimulus class as the native words, at least in early phases of learning. Research on stimulus equivalence has repeatedly shown that following the establishment of at least two overlapping conditional discriminations, new and untrained relations emerge (see for example Sidman, 1971, 1992; Sidman, 1994; Sidman & Tailby, 1982). According to Sidman, this generative outcome is one of the most interesting features of stimulus equivalence. In Sidman's own word "a small amount of teaching can yield a tremendous amount of learning" (Sidman, 2009, p. 14). This feature has begun to receive more attention recently in classroom applications, as stimulus equivalence procedures lead to the emergence of a great number of untaught relations, based on a rather small amount of directly trained relations (see Arntzen, 2010; Arntzen, 2012). Although the effect of a variety of procedural variables and arrangement has been investigated in both laboratory and applied research, the key feature of EBI involves teaching at least two overlapping conditional discriminations and then testing for emergent responding. For example, following the establishment of AB and BC conditional discriminations, test for emergent responding will include probes for BA and CB relations (symmetry property), AC relations (transitivity property), and CA relations (equivalence property).

In recent years, Equivalence-Based Instruction (EBI) has been used to teach a variety of skills (see Pilgrim, 2020, for an overview). In one study, adolescents with autism have been successfully taught novel academic skills in topics like mathematics, history, and science (Stanley et al., 2018). Theoretical music skills and piano playing have been taught in EBI to female college students (Griffith et al., 2018) and children with autism (Arntzen et al., 2010). Preschool children with autism learned how to categorize toys, clothing, and fruits in another study (Barron et al., 2018). Multiple studies have demonstrated the success of EBI on teaching different skills to college students, such as behavior analytic concepts (Augland et al., 2020), statistical variability (Albright et al., 2015), accurate estimation of different portion sizes (Trucil et al., 2015), identifying logical fallacies (Ong et al., 2018), and interpreting operant functions of behavior (Albright et al., 2016). EBI seems to be a promising method with regards to teaching concepts in both children and adults.

While stimulus equivalence has not been intensively investigated when it comes to teaching second language, there are some studies that illustrate how EBI could work in teaching health related concepts in a classroom setting. For example, a study by Albright et al. (2016) taught interpretation of operant functions of behavior to graduate students. EBI readily established baseline relations and demonstrated equivalence class formation. Furthermore, students' performance on written post-tests improved significantly after EBI. Similarly, an EBI study by Fienup et al. (2010) demonstrated that 64 untrained stimulus-stimulus relations emerged following training of 16 baseline relations. Furthermore, medicine students in a study by Greville et al. (2016) rated the EBI experience as positive and three parts of the student preferred the EBI format to textbooks.

Stimulus equivalence has not been used extensively in the research of second language acquisition; furthermore, the generative power of stimulus equivalence procedures in teaching second language vocabularies has been largely ignored (Sidman, 2009). The first study done in this area, to our knowledge, involves teaching Spanish words to English speaking adolescents who had previous traumatic head injuries. The results showed that the despite receiving training on only one task, the participants showed emergence

of new relations, as well as retention of the new relations after a maintenance test (Joyce, 1993). Stimulus equivalence procedures have also proven effective in teaching English speaking preschool children's numerals in Dakota and Ojibwe, two native American endangered languages (Haegele et al., 2011). The study involved an intervention group, in which participants received a computerized matching-to-sample procedure and a control group, in which participants received usual classroom instruction. The results showed that all the participants who received training on baseline relations developed equivalent classes, thus demonstrating that EBI may be used successfully in teaching second language words and numerals.

In another study, one English speaking child was taught conditional discriminations between Spanish words and pictures, while her brother observed the procedure (Ramirez et al., 2009). Symmetry relations emerged in both participants during the posttest despite only one of them being directly taught dictated names in Spanish. The conditional-discrimination training proved to be effective in the acquisition of oral naming skills, as well as it demonstrated that observation could lead to the emergence of derived symmetry relations. Stimulus equivalence-based procedures such as MTS and constructed-response-matching-to-sample, were utilized in the acquisitions and transfer of English as second language for Japanese students with autism. The students developed both English writing and reading skills successfully, implying stimulus equivalence procedures might yield effective results for individuals with cognitive disabilities (Omori, 2011).

While there is a general agreement that immigrants should and must learn Norwegian to succeed, there are several approaches when it comes to how much the native language is used in second language instruction. In a study, which spanned from 2011 to 2014, the native languages of the students were used actively when teaching Norwegian to adults with foreign background in public institutions (Alver & Dregelid, 2016). The results of this interview-based study showed that the participants expressed improvement when it came to understanding keylanguage concepts, as well as a boost in confidence and motivation. The effectiveness of using the native language of the immigrants when teaching the host country language is also supported by the concept of *translanguaging*, which is described as a "tool to mediate complex social and cognitive activities" (Garcia & Wei, 2014, p. 137). Translanguaging is a linguistic concept that supports the use of different languages together to improve communication. As commented above, there is a pronounced need for more effective teaching formats in second language acquisition. Furthermore, investigations of teaching components such as the use of native language in stimulus materials can potentially inform future instructional design. Hence, the present study also incorporates the use of the native language as a variable with the purpose of finding out how it might impact the learning process.

Based on the limited, but promising body of research involving stimulus equivalence and second language acquisition, the purpose of the present study was to explore the effects of EBI in teaching Norwegian health related concepts to individuals with different native languages. Furthermore, we wanted to investigate the relative effects of arranging stimulus sets with Norwegian text only and stimulus sets with some stimuli in the participants native language on equivalence class formation.

Method

Participants

Eight adult participants with a minority background, age between 32 and 63 years (average 39), were voluntarily recruited for this study. All participants were attending Norwegian language classes for adults at a public teaching institution. In addition to

learn Norwegian grammar and syntax, all participants were also attending classes to further their education in health sciences at the high school level. The Norwegian language skills level of the participants varied from A1 to B1, according to a scale used by the Common European Framework of Reference for Language (Kompetanse-Norge, 2017), where A levels are beginners and B levels are intermediate.

Setting and Material

The experiment took place in the students' ordinary classroom. Workstations were placed in a U shaped classroom layout, with 23 computers and chairs. The participants were asked to turn off their phones and instructed to ask the experimenter if they needed help or required a break. The time frame for the experiment was between 9 AM and 1:30 PM, meaning the participants had approximately 4,5 hours to complete the experiment. MTS training and testing were run on Lenovo Think Center stationary computers with Windows 10 Enterprise operating system and 24-inch screens. The project was approved by the Norwegian Center for Research Data (NSD). The MTS software was customized for the tasks and automatically registered data such as the trial types, blocks, number of correct and incorrect comparison choices and programmed consequences. A joint debriefing session was announced at start-up and offered at the end of the experiment.

Design

We arranged a combined within- and between-subject design, in which each participant was tested for equivalence class formation subsequent to MTS training of conditional discriminations. Participants were quasi-randomly assigned to one out of two experimental conditions: (a) all text stimuli in Norwegian and (b) some text stimuli in native language. The native languages available were Tamil, Somali, and Tigrinya.

Independent and Dependent Variables

The independent variables were MTS training of overlapping conditional discriminations with D-stimuli in different languages and pictures. The dependent variable was performance in MTS training, measured as number of trials to criterion and equivalence class formation.

Procedure

Instructions

The instructions for the experiment were given while the participants sat at the desks without the computers, and then they were instructed to sit at the computer desks before proceeding with the experiment. Two experimenters were in the classroom; one positioned at the back and one at the front. The experimenter at the front provided oral instructions at the onset of the experiment. One teacher, familiar with the language level of each participant, was present in the classroom to assist with repeated instruction if needed. The instructions were visually aided by demonstrations of identity matching (colors) and arbitrary matching tasks, which were presented via Power Point on a white board screen. The stimuli in this sequence were different from the ones used in the experiment. In the identity matching task, the stimuli were comprised of blue, green, orange, and yellow circles. The arbitrary matching exercise featured text and picture stimuli of dog, cat, horse, and cow. Both identity and arbitrary matching demonstrations included (a) response to sample, (b) a correct comparison choice with programmed consequence, (c) an incorrect comparison choice with programmed consequence. As an additional measure to mimic the tasks featured in the experiment, the position of the alternatives switched for every sequence. Towards the end of the instructional demonstrations the participants were encouraged orally to: "Do your best to answer every task correctly".

Stimuli

The stimulus set was based on the curriculum in the health science course. The

participants were familiar with concepts similar to the ones featured in the stimulus sets. Furthermore, the definitions and concepts used in the stimulus set were also cross referenced with medical dictionaries such as Store Medisinske Leksikon (Store medisinske leksikon, n.d.), used by many Norwegian universities. The languages selected for the experiment were the ones most represented among participants and where translation also was available.

Figure 1 gives an overview of the stimulus sets used in the experiment. There are four potential classes marked with the numbers 1, 2, 3, and 4 on the first row, while the four members of each class are presented vertically and are marked with A, B, $C_{1, 2, 3}$ and D. The A-stimuli are names of four main systems of the body: A1 is the skeletal system, A2 is the cardiovascular system, A3 is the digestive system, and A4 is the system comprised of muscles and joints. The B-stimuli were either black and white or colored pictures of the four systems. The C_1-stimuli were comprised of definitions or descriptions of the four systems and were approximately equal in length, ranging from nine to 12 words per definition. The C_2- and C_3- variants were descriptions of a specific component of the relevant body systems. For example, $C3_2$ was stomach and $C3_3$ was small intestine. The three variations of C-stimuli were arranged to avoid undesired stimulus control by the shape of the text. Since one group experienced Norwegian text only and the other group text in their native language, the D-stimuli were different for the two groups. D-stimuli for the Norwegian text only group were a close-up or more detailed picture of the B-stimuli. For the native language group, the D-stimuli represent the A-stimuli translated into their respective native language. The translation of the concepts in these respective foreign languages was done by professional translators, and the accuracy of the translation was revised and verified by native speakers of the three languages.

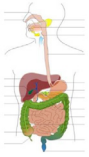
All stimuli were displayed on white background and the written stimuli were presented in black letters. The size of the stimuli was approximately 5 x 6 centimeters, implying that a mouse click within those frames would produce the next element of the trial.

MTS Training

Figure 2 shows an overview of the procedure. Twelve conditional discriminations were trained in a one-to-many (OTM) training structure where A-stimuli served as sample stimuli. We arranged a simultaneous training and testing protocol, in which all baseline discriminations were established to criterion before test for emergent relations. Baseline relations were introduced in a serialized manner, implying that AB relations were established to mastery criterion before AC relations were introduced and established to mastery criterion. AC training was followed by a mixed phase of AB and AC relations, before AD relations were introduced and established to mastery criterion.

Each trial started with the presentation of the sample stimulus in the center of the screen. When the participant clicked on the sample stimulus, four comparison stimuli appeared in the corners of the screen. The position of comparison stimuli varied from trial to trial. If the participant clicked on the corresponding comparison stimulus, the Norwegian word for correct, excellent, and nice appeared in the center of the screen in blue letters on white background. If the participant selected a non-corresponding stimulus, the Norwegian word for wrong appeared in the center of the screen. Each trial was followed by a 500 ms inter-trial interval.

Each baseline relation was presented six times in a block, implying that each variant of C-stimuli was presented two times. Within each block, the different baseline relations were presented in random order. The criterion to proceed to the next training phase was a minimum of 95% correct responding. If a participant did not reach the criterion, the training block was repeated.

	1	2	3	4
A	Skjelettet	Hjerte-karssytemet	Fordøyelsessystemet	Muskler og ledd
B				
C_1	Holder kroppen oppreist, består av knokler, beskytter indre organer	Organsystem som frakter blod rundt i kroppen; temperaturregulering og transport av stoffer	Kanal som maten må gå gjennom før kroppen får næringsstoff	Vev som kan trekkes sammen og struktur som binder to knokler sammen
C_2	Lårben	Arterier	Magesekk	Bindevevsforbindelser
C_3	Ryggsøylen	Vener	Tynntarm	Glatt muskulatur

D Native Language

Tamil	எலும்பு மண்டலம்	இதயகுழலிய மண்டலம்	செரிமான மண்டலம்	மூட்டுகள் மற்றும் தசைகள்
Somali	nidaamka lafaha	nidaamka wadnaha	nidaamka dheef-shiidka	kala-goysada iyo muruqyada
Tigrinja	እቲ ስርዓተ ኣስከሬን	ስርዓተ መትነ-ልባዊ	እቲ ስራዓተ ምሕቃቕ ምግቢ	መላግቦታትን ጭዋዳታትን

D Norwegian only

Figure 1. Stimulus Set.

Thinning of Programmed Consequences. When mastery criterion was reached for mixed phase with all conditional discriminations (72 trials per block), participants experienced three training phases with gradual thinning of programmed consequences. The thinning steps were 75%, 50%, and 0% probability of consequences. Mastery criterion was the same as previous training phases.

Test for Stimulus Equivalence

The test for equivalence class formation followed immediately after last training phase and included all emergent relations with probes for maintenance of baseline relations interspersed. Each trial type was

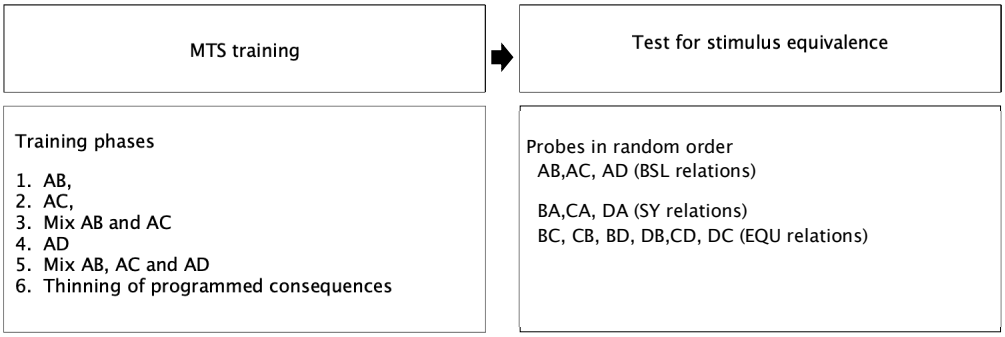


Figure 2. Overview of the Procedure.
Note: This figure illustrates the simultaneous protocol in the upper row, the lower left box describes serialized introduction of baseline trials and lower right box shows probes presented in the test. BSL =baseline, SY=symmetry and EQU= equivalence.

presented six times during the test, constituting a total of 216 test trials for emergent relations and 72 maintenance probes for baseline relations. All trial types were presented in random order and without programmed consequences. Symmetry trials included BA, C_1A, C_2A, C_3A, and DA. BA and DA trials were presented six times each, while C_1A, C_2A, and C_3A were presented two times each, yielding a total of six CA relations. Equivalence trials included BC, CB, BD, DB, CD, and DC trials, with all three C variants. BD and DB probes were presented six times, while the three variants of C-stimuli in were presented twice for each of the BC, CB, CD, and DC probes. The experimenter-defined criterion to conclude that an equivalence class was formed was ≥ .9 for baseline, symmetry, and equivalence trial types, respectively.

Results

Participant Flow

Eight participants, four in each group, completed the experiment. A total of twelve participants were originally recruited. Two participants did not finish within the predetermined time frame (4.5 hours), and one withdrew mid-experiment due to worsened eyesight caused by sitting in front of a PC for long periods of time.

Number of Training Trials

Table 1 shows the individual results for all participants. Due to a programming error, P18205 experienced each trial type only one time per training block; thus, the number of trials required to reach the criterion is reported as number of trials above the minimum number of trials required by the procedure. On average, participants experiencing D-stimuli in their native language required fewer trials to criterion (M198, range 40–448), than did participants experiencing text stimuli in Norwegian only (M=358, range 40–640).

Equivalence Class Formation

In total, seven out of eight participants formed equivalence classes. Individual results are shown in Table 1, where column BSL, SY and EQU mark the percentage of correct responding for maintenance of baseline trials, symmetry, and equivalence, respectively. The equivalence criterion was ≥.9 for each of the trial types.

Among participants who experienced Norwegian words only, three out of four participants formed classes. P18202, P18204, and P18214 reached criteria for baseline, symmetry, and equivalence, as they scored above the required .9 criterion. P18212 demonstrated emergent symmetry relations but scored just below the .9 benchmark on equivalence trials.

Table 1. Participant characteristics and individual results.

Condition	Participant #	Trials > Min.	BSL	SY	EQU
Native	18211	40	**99**	**100**	**100**
	18205*	152	N/A	**100**	**100**
	18207	136	**99**	**100**	**99**
	18203	448	**96**	**96**	**94**
Norwegian Only	18204	40	**99**	**100**	**100**
	18214	112	**96**	**100**	**100**
	18202	640	**100**	**99**	**99**
	18212	640	**94**	**93**	89

Note: BSL = percentage correct for baseline relations. SY = percentage correct for symmetry relations, EQU = percentage correct for equivalence relations, and Trials > Min = trials above minimum. Numbers in bold marks that the equivalence criterion was reached. Asterisk denotes that due to a programming error the participant experienced each trial type only one time per training block and test without probes for maintenance of baseline relations.

All participants who experienced native language as D-stimuli formed equivalence classes. For one participant, P18205, the script was not properly loaded at onset of the experiment. Due to this programming error, this participant did not experience maintenance probes for baseline relations during test conditions. Still, the high outcome on emergent probes suggests that baseline relations were intact.

Error Analyses

To get a more detailed picture of performance, we analyzed error patterns in test. P18203 and P18212 experiencing D-stimuli in the native language and text stimuli in Norwegian only, respectively, had more than one incorrect for some trial types. Their error analyses are displayed in Table 2. The table shows that P18203 made two incorrect comparison choices for three trial types, of which three were in experimenter defined class four and one in experimenter defined class two. Furthermore, two of the trial types involved DC/CD relations. P18212 made 25 incorrect comparison selections in test, of which ten were in experimenter defined class one and 15 in class four. Despite a relatively high number of negative test probes, no errors occurred for class two and three.

Discussion

The purpose of this study was to investigate the effectiveness of EBI in teaching health related concepts in Norwegian as a second language to adults with a foreign background. Furthermore, we also wanted to explore whether using the native language of the participants would impact positively the learning process.

Table 2. Test Performance with More than 1 Incorrect per Trial Type.

Participant	Total Errors	>1 Error BSL	>1 error SY	>1 errors EQU
18203	15	2x A4C4		2x **C2D2**, B4C4, **D4C4**
18212	25	2x A1C1 and A4C4	2x C1A1, C4A4	3x **C4D4**; 2x B1C1, **C1D1, D1C1**, B4C4, C4B4, **C4D4** and **D4C4**

Note: BSL=baseline relations, SY= symmetry, EQU=equivalence and > = more than. Trial types in bold involve D-stimuli.

Equivalence Class Formation

Overall, the results showed that seven out of the eight participants who completed the experiment demonstrated equivalence class formation, while one of them reached the criterion for baseline and symmetry relations, barely missing the criterion for equivalence relations. In other words, with the current procedure seven out of eight participants demonstrated a simple understanding of Norwegian health related concepts they had previously struggled to learn. These results support that equivalence-based technology are effective when teaching a second language, as suggested by the study by Haegele et al. (2011), in which English-speaking preschool children were taught numerals in Dakota and Ojibwe Studies involving stimulus equivalence and second language learning have not been abundant and have not targeted typically developed adults. The current results build on previous research where autistic Japanese students were taught English (Omori, 2011), as well as a study where English-speaking head injured adolescents were taught Spanish (Joyce, 1993).

The use of Native Language in Second Language Acquisition

Pertaining to the question of whether involving the native language for some of the participants would improve their performance, we failed to obtain a differential outcome between the two conditions. As such, the present results differ from studies focused on bilingualism conducted in professions other than behavior analysis, where the use of native language in second language acquisition impacted the learning outcomes positively. For example, a study by Salmona Madriñan (2014) showed that the students were able to understand complex concepts in their native language, and knowledge was transferred to a second language (English). The study consisted of a curriculum where English was the only language used when providing instructions, and eventually another curriculum was introduced where both Spanish and English were used. The Salmona Madriñan study presents similarities to the present study; in that it has one condition involving only the target language and another condition involving a mix of the second language and the native language. It also differs by having a greater number of participants, and by that it focuses on comprehension and concept transfer. Another study conducted with adults learning English with Arabic as their first language has shown that the use of Arabic facilitated the understanding of vocabulary and expressions in English (Khan, 2016). The results of this study also suggest that the students who received bilingual training remembered more words and expressions overtime compared to students experiencing English-only instruction. A facilitating effect of familiar stimuli are also supported by basic research in stimulus equivalence where studies have shown that stimulus sets with familiar stimuli results in higher yields of equivalence class formation than abstract stimuli only (e.g. Arntzen & Mensah, 2020; Arntzen et al., 2018; Hayashi & Vaidya, 2017).

A low number of participants by itself can make a differentiated outcome hard to obtain and another possible source for the disparity between present results and that of others is the stimuli used as D-stimuli. In the present study D-stimuli for the native group was text in their mother's tongue, while the Norwegian group were pictures showing a part of, or a close-up, of the pictures used as B-stimuli. It is possible that close-up pictures are just as "familiar" as text in the native language and as such has equalized the conditions. To explore this possibility, future studies should compare text in native language with text in host country language.

A third procedural detail that should be discussed is that the A sample stimuli during the training session were the written Norwegian words for the four main anatomical body systems. If using the native language

makes a difference in second language acquisition, it is possible to arrange (in the one-to-many, or OTM, training structure) the native language words as A-stimuli; thus, differentiating the two conditions in a more visible way. OTM training structures present the A-stimuli for each trial, meaning that with four classes, each A-stimulus appears roughly every fourth trial, which is more often than the comparison stimuli. Thus, future should consider using native language words as sample in OTM training structures and as comparison stimuli in MTO training structure. This would make the native language stimuli appear more frequently during training and might contribute to a more conclusive result with regards to the role of using native language in teaching a second language.

Error Analyses

The error analysis of participants with most errors in test, P18203 and P18212, show that relatively many errors occurred with stimulus relations involving C-stimuli. For both groups C-stimuli involved descriptions in Norwegian and to avoid irrelevant stimulus control by for example form of text, we also arranged three variants of each of the stimuli. This means that each of the C-stimuli alternatives were presented only twice in each block, as opposed to six presentations of B and D stimuli. It is possible that this arrangement influenced the negative test scores. Due to a programming error and contrary to the results of these two participants, P18205, experienced each trial type only once in each block. However, this participant experienced each of the C-variants in separate blocks and had 100 per cent correct on probes for symmetry and equivalence. It is possible that the difference in baseline training has influenced the disparate outcome in these participants. Thus, future studies should explore this possibility. We also noticed that that both groups during training struggled with the same types of trial, namely A4-C4 and especially variant 2 of the C-stimuli, which was the word *bindevevsforbindelser*. This could be explained by the novelty of the word, perhaps a less common concept present in their curriculum. The potential influence of this variable should be explored in future experiments, where the length of some text stimuli and their novelty is manipulated.

Limitations and further research

There are some limitations with the present study. First, the present study was restricted to the availability of students, and the necessary time required to complete the study. Due to practical circumstances and voluntary participation, the students had to complete the experiment within the allocated time frame and four participants did not finish the experiment. The present experiment arranged four potential classes with four members per class and in addition there were three variants of the C-stimuli with relatively complex and lengthy words. One way to avoid this problem in future studies could be to adjust the number and size of the classes targeted to the allocated time frame.

A second limitation is the design used to evaluate within-subject changes in responding. The present study did not arrange a pre-test or a retention test due to time restrictions. A pre-test could have potentially excluded pre-experimental histories as an alternative explanation for test performance and reduce the influence of threats to internal validity. A retention test could potentially contribute to both the external and the social validity of the study. One of the appeals of stimulus equivalence procedures is that concepts are taught economically, but it would be even better if one could demonstrate that the concepts acquired would maintain over time. Thus, future replications should include both design elements.

Finally, the present experiment featured Tigrinya, Somali and Tamil as the native languages included in the experiment.

One disadvantage of using several native languages stems from the uncertainty on the proficiency of the participants in their native language when it comes to medical concepts. In addition, the languages chosen were based on how many speakers of the respective language were in that student group. That means that not all the languages spoken in the student group were featured in the experiment. From the applied perspective then, we were not able to offer EBI to all students in the group. Adding variation when it comes to the participants' native languages could also contribute to demonstrating the generality of the procedure and its applicability across several languages spoken by participants who are learning Norwegian. This is an option that should be explored in future experiments. The present study expands previous demonstrations by having typically developed adults as participants, as well as Norwegian language as the target language with several native languages like Tigrinya, Somali, and Tamil.

Despite these limitations, the present study shows that EBI readily formed equivalence classes in seven out of eight participants. Even though stimulus equivalence has been used in teaching concepts to adults, there are very few experiments involving second language acquisition and stimulus equivalence. The present study also capitalizes on the present political and social environment regarding immigration policies and the importance of learning Norwegian as a part of the integration processes. In addition, the study was customized for participants who attend adult learning institutions and who are in the processes of learning both the language and healthcare curriculum, to gain better access to the jobs. That takes stimulus equivalence from laboratory conditions into a real-life scenario where there are potential direct benefits for the participants. This could also increase the appeal of stimulus equivalence procedures being integrated into traditional learning environments. This type of procedure could be implemented as a revision tool before exams, especially in relation to difficult concepts students usually learn by heart. What could also be relevant in promoting the use of stimulus equivalence procedures in a classroom setting, is that the teachers learn more about the efficiency and economy of EBI, where many untaught relations can emerge from teaching a handful of relations (Stromer et al., 1992).

Concluding remarks

Seven out of eight participants formed equivalence classes involving Norwegian health related concepts. The potential of using stimulus equivalence procedures in second language acquisitions has not been fully explored and the number of studies that inquire into the effects of stimulus equivalence on second language learning is small. However, the studies show promising results and present results adds to these findings. On the issue of whether using the native language of the participants in the instructional format would facilitate learning, the results are inconclusive. Future replications should explore potential effects of arranging native language stimuli as node and compare this to a condition where the host language serve the same role.

Compliance with Ethical Standards and Data Availability Statement

Conflict of Interest: The authors declare that there is no conflict of interest.

Ethical Approval: All procedures performed in studies involving human participants were in accordance with the ethical standards and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The study was approved by the Norwegian centre for research data.

Informed Consent: Informed consent was obtained from all individual participants included in the study.

Data Availability Statement: De-identified data are available from the second author.

References

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

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

Alver, V., & Dregelid, K. M. (2016). «Vi kan lære som vanlige folk» - Morsmålsstøttet undervisning. *Bedre Skole*. https://utdanningsforskning.no/artikler/vi-kan-lare-som-vanlige-folk---morsmalsstottet-undervisning/

Arntzen, E. (2010). Stimulusekvivalens. Teoretiske betraktninger og noen praktiske implikasjoner. In F. Svartdal & S. Eikeseth (Eds.), *Anvendt Atferdsanalyse (2. Utg.)* (pp. 100–138). Gyldendal Akademika.

Arntzen, E. (2012). Training and testing parameters in formation of stimulus equivalence: Methodological issues. *European Journal of Behavior Analysis, 13*, 123–135. https://doi.org/10.1080/15021149.2012.11434412

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

Arntzen, E., & Mensah, J. (2020). On the effectiveness of including meaningful pictures in the formation of equivalence classes. *Journal of the Experimental Analysis of Behavior*, 113, 305–321. https://doi.org/10.1002/jeab.579

Arntzen, E., & Nartey, R. K. (2018). Equivalence class formation as a function of preliminary training with pictorial stimuli. *Journal of the Experimental Analysis of Behavior, 110*(2), 275–291. https://doi.org/10.1002/jeab.466

Arntzen, E., Nartey, R. K., & Fields, L. (2018). Graded delay, enhanced equivalence class formation, and meaning. *The Psychological Record*, 68, 123–140. http://doi.org/10.1007/s40732-018-0271-6

Augland, H., Lian, T., & Arntzen, E. (2020). Comparing a student active learning format to equivalence-based instruction. *European Journal of Behavior Analysis, 21*(2), 328–347. https://doi.org/10.1080/15021149.2020.1752513

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

Djuve, A. B., Kavli, H. C., Sterri, E. B., & Bråten, B. (2017). *Introduksjonsprogram og norskopplæring: Hva virker for hvem?* https://www.fafo.no/images/pub/2017/20639.pdf

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

Garcia, O., & Wei, L. (2014). *Translanguaging: language, bilingualism and education* (1st ed.). Palgrave Macmillan UK: Imprint: Palgrave Pivot.

Greville, W. J., Dymond, S., Newton, P. M. (2016). The student experience of applied equivalence-based instruction for neuroanatomy teaching. *Journal of Educational Evaluation for Health Professions, 13*, 32–32. https://doi.org/10.3352/jeehp/2016.12.32

Griffith, K. R., Ramos, A. L., Hill, K. E., & Miguel, C. F. (2018). Using equivalence-based instruction to teach piano skills to college students. *Journal of Applied*

Behavior Analysis, 51(2), 207–219. https://doi.org/10.1002/jaba.438

Haegele, K., McComas, J., Dixon, M., & Burns, M. (2011). Using a stimulus equivalence paradigm to teach numerals, English words, and Native American words to preschool-age children. *Journal of Behavioral Education, 20*(4), 283–296. https://doi.org/10.1007/s10864-011-9134-9

Hayashi, Y., & Vaidya, M. (2017). Effects of stimulus meaningfulness on the acquisition ofconditional relations and the emergence of symmetrical relations in adult humans. *European Journal of Behavior Analysis, 18*(1), 99–112. https://doi.org/10.1080/15021149.2016.1264147

Joyce, B. G. (1993). Using stimulus equivalence procedures to teach relationships between English and Spanish words. *Education and Treatment of Children, 16*(1), 48–65.

Kaida, L. (2013). Do host country education and language training help recent immigrants exit poverty? *Social Science Research, 42*(3), 726–741. https://doi.org/10.1016/j.ssresearch.2013.01.004

Khan, M. S. (2016). The impact of native language use on second language vocabulary learning by Saudi EFL students. *English Language Teaching, 9*(5), 134–140. http://dx.doi.org/10.5539/elt.v9n5p134

Kompetanse-Norge. (2017). Norskkrav i arbeidslivet. https://www.kompetansenorge.no/Norsk-og-samfunnskunnskap/Norskprove/norskkrav-i-arbeidslivet/Pob=12717

Kunnskapsdepartementet. (2019). Opplæring i norsk og samfunnskunnskap. https://www.regjeringen.no/no/tema/innvandring-og-integrering/kd/Verkemiddel-i-integreringsarbeidet/opplaring-i-norsk/id2343471/

Omori, M. (2011). Acquisition and transfer of English as a second language through the constructional response matching-to-sample procedure for students with developmental disabilities. *Psychology, 02*, 552–559. https://doi.org/10.4236/psych.2011.26085

Ong, T., Normand, M. P., & Schenk, M. J. (2018). Using equivalence-based instruction to teach college students to identify logical fallacies. *Behavioral Interventions, 33*(2),22–135. https://doi.org/10.1002/bin.1512

Pilgrim, C. (2020). Equivalence-based instruction. In J. O. Cooper, T. E. Heron, & W. L. Heward (Eds.), *Applied Behavior Analysis* (3rd ed., s. 452–496). Pearson Education. Ramirez, J., Rehfeldt, R. A., & Ninness, C. (2009). Observational learning and the emergence of symmetry relations in teaching spanish vocabulary words to typically developing children. *Journal of Applied Behavior Analysis, 42*(4), 801–805. https://doi.org/10.1901/jaba.2009.42-801

Salmona Madriñan, M. (2014). The use of first language in the second-language classroom: Asupport for second language acquisition. *GIST Education and Learning Research Journal*, *9*, 50–66. https://doi.org/10.26817/16925777.143

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

Sidman, M. (1992). Equivalence relations: Some basic considerations. In *Understanding verbal relations* (pp. 15–27). Context Press.

Sidman, M. (1994). *Equivalence relations and behavior: A research story*. Authors Cooperative.

Sidman, M. (2009). Equivalence relations and behavior: An introductory tutorial. *Analysis of Verbal Behavior, 25*, 5–17. https://doi-org.ezproxy.hioa.no/10.1007/BF03393066

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

Stanley, C. R., Belisle, J., & Dixon, M. R. (2018). Equivalence-based instruction of academicskills: Application to adolescents with autism. *Journal of Applied Behavior Analysis, 51*(2), 352–359. https://doi.org/10.1002/jaba.446

Store medisinske leksikon. (n.d.). https://sml.snl.no/

Stromer, R., Mackay, H., & Stoddard, L. (1992). Classroom applications of stimulus equivalence technology. *Journal of Behavioral Education, 2*(3), 225–256. https://doi.org/10.1007/BF00948817

Trucil, L., Vladescu, J., Reeve, K., DeBar, R., & Schnell, L. (2015). Improving portion-size estimation using equivalence-based instruction. *The Psychological Record, 65*(4), 761–770. https://doi.org/10.1007/s40732-015-0146-z