CORE
Metadata, citation and similar papers at core.ac.uk
Provided by IUScholarWorks

INDIANA UNIVERSITY
BLOOMINGTON

Gierut / Learnability Project

Learnability Project

The Learnability Project was founded in 1985 by Judith A. Gierut, Professor Emerita of Speech and Hearing Sciences, Indiana University. Through funding from the National Institutes of Health, the project served as a test site in evaluation of the efficacy of clinical treatment for preschool children with functional (nonorganic) phonological disorders. Children who enrolled contributed longitudinal descriptive phonological samples for linguistic analysis. They also received clinical treatment, designed as single-subject experiments, to establish the optimal teaching conditions to promote phonological learning. Experimental studies were based on the triangulation of theoretical models of linguistics, psycholinguistics, and speech-language pathology, with the aim of bridging theory with application and science with best practices. The Gierut / Learnability Project collections accord with the data-sharing plan of the National Institutes of Health and are intended for broad use by scientists, clinicians, and students interested in language and learning.

Content Statement

This document was retrieved from the BASICS collection of the Gierut / Learnability Project archived in IUScholarWorks https://scholarworks.iu.edu/dspace/handle/2022/20061. By downloading and/or using the content herein, the user agrees to the Creative Commons copyright license CC BY-NC-ND. The copyright license permits the user to access and share the Gierut / Learnability Project collections, with appropriate acknowledgement and credit to the creators/authors. The copyright license denies the user the right to make changes to the collections or develop the collections commercially. The terms of the copyright license deed may be found at http://creativecommons.org/licenses/by-nc-nd/4.0/ and the legal code at http://creativecommons.org/licenses/by-nc-nd/4.0/legalcode.

Learnability Project Working Paper

Experimental Designs and Protocols

Judith A. Gierut
Indiana University

Learnability Project working papers were developed for internal purposes in the training of research assistants. The material herein first appeared in the Learnability Project Lab Manual, version 1, 1986, and was updated as the protocol expanded. This working paper outlines, in part, some general principles of single-subject experimental design and summarizes some core elements of the treatment and generalization protocols used in Learnability Project research. It is the companion to the Experimental Archive of the DATA collection of the Gierut / Learnability Project. The working paper is not intended as a comprehensive review of single-subject experimental design or the range of experimental designs and protocols used by the Learnability Project. The reader is referred to primary source material found in the Publications collection of the Gierut / Learnability Project archived in the IUScholarWorks repository. The following texts and publications may be particularly useful as introductions to the population, experimental design, and lab protocols.

Suggested Readings:

Gast, D. L. (Ed.) (2010). *Single subject research methodology in behavioral sciences*. New York: Routledge.

Gierut, J. A. (1998). Treatment efficacy: Functional phonological disorders in children. *Journal of Speech, Language and Hearing Research*, *41*, S85-S100. PMID: 9493748

Gierut, J. A. (2005). Phonological intervention: The how or the what? In A. Kamhi & K. Pollack (Eds.), *Phonological disorders in children: Clinical decision making in assessment and intervention* (pp. 201-210). Philadelphia: Brookes.

Gierut, J. A. (2008a). Fundamentals of experimental design and treatment. In D. A. Dinnsen & J. A. Gierut (Eds.), *Optimality theory, phonological acquisition and disorders* (pp. 93-118). London: Equinox.

Gierut, J. A. (2008b). Phonological disorders and the developmental phonology archive. In D. A. Dinnsen & J. A. Gierut (Eds.), *Optimality theory, phonological acquisition and disorders* (pp. 37-92). London: Equinox.

Gierut, J. A., Morrisette, M. L., & Ziemer, S. M. (2010). Nonwords and generalization in children with phonological disorders. *American Journal of Speech-Language Pathology*, *19*, 167-177. PMCID: PMC3281489

McReynolds, L. V., & Kearns, K. P. (1983). *Single-subject experimental designs in communicative disorders*. Baltimore, MD: University Park Press.

 ©Judith A. Gierut, 2015

Introduction to Single-Subject Design

Single-subject design is a research approach that was initiated in the 1960's as part of the behaviorist movement, but has gone well beyond the narrow scope of its founding. It is relevant to the evaluation of learning through instruction and applicable to a wide range of populations. It is well suited to the study of development because an understanding of the trajectory of change is one of its outcomes. It is also a good match to the study of phonological disorders because these children require treatment to bring their sound systems in line with the target phonology. Like group design, single-subject design has all of the makings of an experimental paradigm including (1) operationally defined independent and dependent variables, (2) control of extraneous variables, (3) data that are reliable and valid, and (4) generalizability to the population at-large. However, the way single-subject design achieves these criteria is best understood by its underlying assumptions.

By its name, single-subject design emphasizes the individual. There are few participants in any given study. Contrary to the assumption of homogeneity in between-subjects designs, single-subject design assumes heterogeneity of the population. This fits with the study of phonology, acquisition, and disorders given the range of variation that occurs within and across languages. For children with phonological disorders, heterogeneity is the norm rather than the exception.

Independent and Dependent Variables

The independent variable is the property that is manipulated in an experiment, and it is determined a priori by the hypothesis being tested. The dependent variable is the property that is measured in an experiment, and it too is dictated by the hypothesis. In single-subject design, the dependent variable is typically a measure of generalization learning. *Generalization* is the transfer of a newly acquired property in treatment to other untreated and previously unknown properties. Generalization refers to the extension of learning that benefits or promotes gains in behavior following a minimum of instruction. Three types of generalization are often cited as dependent variables in studies of phonology.

Generalization to the treated class refers to the transfer of the treated sound to untreated words and contexts. For example, if /f/ were taught in word-initial position, generalization would be measured in word-initial, intervocalic, and final contexts using novel stimuli. Within-class generalization is the transfer of learning to untreated properties that are related to the treated property. In the example of word-initial /f/ as the treated sound, within-class generalization would be monitored for other untreated fricatives in error. Across-class generalization refers to the transfer of learning to untreated properties that are not related to the treated property. For the case of /f/, across-class generalization might be sampled in liquids or affricates if these classes were in error. It is important to note that the notion of a generalization "class" or category is defined by the variables of study. In the example of /f/, treated, within- and across-class generalization was defined along the dimension of manner. However, it is equally possible to define the generalization class along any number of linguistic, psycholinguistic, or clinical variables, e.g., distinctive features, syntactic category, word frequency, normative age of sound mastery. Category membership is delineated by the goal of treatment (i.e., the independent variable). If the goal is to teach a distinctive feature contrast, then features become the generalization class. Similarly, if teaching is aimed at verbs, then syntactic category is the relevant generalization variable. If teaching emphasizes early-acquired words, then normative age defines the generalization sample. An ultimate goal of phonological treatment is to induce system-wide gains in production accuracy that extend across sounds, contexts, and words.

 ©Judith A. Gierut, 2015

Design Assumptions

Inherent in the set up of a single-subject experiment is that premise that every child serves as his or her own demonstration of experimental control. The demonstration of control means that there is a causal relationship established between the variables of study, which cannot be attributed to other extraneous factors. In most applications of single-subject design, the cause-effect involves treatment and learning, such that treatment is responsible for the change in learning. Presumably, the level of “noise” that is associated with extraneous influences in an individual’s daily routine (e.g., fatigue, hunger, illness) is relatively stable, fluctuating within a narrow band. To measure the noise, a repeated baseline of a subject’s performance is obtained prior to the instatement of treatment. This then is factored out of the treatment results. When gains over baseline performance are observed concurrent with, and exclusive to treatment, this establishes a functional relationship between the administration of treatment and learning.

Single-subject design addresses data collection from a similar vantage. Even though only a handful of subjects participate in any given experiment, each contributes multiple points of data. Because the delivery of treatment takes place over time, there is the potential for a micro view of development through trial-by-trial, day-by-day, or step-by-step learning. The quantity of data obtained in a single-subject study often exceeds that collected in one-shot experimental studies. Statistical analyses can be completed using nonparametric time series or effect size analyses; however, the conventional approach to data analysis is through visual inspection of a child’s learning curve, with attention to level and slope of learning.

Single-subject design also stands apart in its handling of the validity of findings. Conventionally, the validity of an experimental finding is confirmed by its generalizability to the population at-large. With few participants in a single-subject design, single-subject design employs direct and systematic replications. A direct replication occurs when children with similar profiles are exposed to the same experimental condition and show similar patterns of learning. A systematic replication takes place when children with different profiles are exposed to different experimental conditions and still show similar patterns of learning. Theoretically, systematic replications are especially valuable because they have the potential to identify higher order principles that govern learning and development.

In all, the properties of single-subject design meet the requisites of an experimental study in the systematic manipulation and measurement of behavior under carefully controlled conditions, with the outcome being replicable, reliable and valid to the extent that it is applicable to a standard normative distribution. The key difference between the single-subject and between-subject models is the emphases on hetero- versus homogeneity respectively. The different vantage points drive the ways in which each model approaches an experimental question.

Fundamentals of the ABA and Multiple Baseline Designs

The ABA is the basic protocol of single-subject design. The ABA consists of three sequential phases. The first A phase is the baseline period, where the dependent variable is measured prior to the application of the independent variable. During baseline, a child’s production of sounds is measured. The baseline phase has three requirements. One is that the behavior is sampled repeatedly over time; another is that the behavior remains stable over the sampling period; and a third is that the same measurement tool is used over the extended period. Repeated and stable baselines demonstrate that a subject’s performance is not in a state of fluctuation or process of change. Baseline stability helps to delineate a child’s entry levels of performance from later performance that follows from the application of the independent

 ©Judith A. Gierut, 2015

variable. It is conventional that two to three baselines be obtained with an interval of time in between. Likewise, the accepted level of fluctuation in baseline performance is $\pm$10% variation. It is also important to examine the trajectory of the baseline, particularly if the trend in performance is rising. A rising baseline is one indicator that the dependent variable is undergoing change in advance of any application of the independent variable. A rising baseline may cloud the interpretation of the effects of the experimental manipulation. Rising baselines are a sign that treatment should not be administered to that behavior being measured.

Following the A phase of repeated stable baseline performance, the B phase of the design is instated. During the B phase, the independent variable is applied, i.e., treatment is administered to affect phonological change. A treatment protocol is developed in advance and may involve a series of graded steps. The B phase typically involves production training, starting with imitative and then advancing to spontaneous outputs. The steps of training continue to pre-established levels, which may be defined by time and/or performance. Time-based criteria for teaching can be set, for example, by number of sessions or trials completed. Performance-based criteria depend on the child achieving a certain level of response accuracy. Time- and performance-based criteria are not mutually exclusive and may be used in conjunction with each other, such that treatment continues for a fixed period of time or until a child reaches a particular level of performance, whichever occurs first. Throughout the B phase, trial-by-trial learning is documented, such that each response is judged as correct or incorrect with appropriate feedback provided to the learner. This is necessary to demonstrate that learning has occurred during treatment and supports that the independent variable has had an observable effect on behavior.

There is a second A phase immediately following completion of treatment, where the baseline measure is re-administered. The assumption is that there will be a return to baseline levels of performance following the removal of the independent variable. This presumably shows that treatment is the factor controlling behavior. When treatment is removed, the behavior returns to pretreatment levels; hence, there is a causal relationship between the independent and dependent variables. The return to baseline in the ABA design is crucial to the demonstration of experimental control; however, it is a point of concern for ecological validity. A return to baseline may be viewed as unethical, particularly where children with phonological disorders are concerned. It is also artificial in the context of language because treatment presumably induces change in a learner's underlying grammar. There are design alternatives to address this concern, with the optimal being use of MBL design in lieu of the ABA. This design is used most often in treatment studies of phonological disorders.

The MBL design involves stacking several ABs in a time-lagged sequence. Multiple ABs provide direct replications of the effects of treatment and serve as a demonstration of experimental control. Each AB replication is called a leg, and the recommendation is that there be at least three AB legs in any given MBL application. Consistent with the AB protocol, the MBL involves a baseline followed by treatment. During the A phase, all legs of the design are measured. As the number of legs increases, so do the number of baselines. The number of baselines to be administered for successive legs of the design is determined by the time-lagged administration of treatment. The assumption of baseline stability still holds to ensure internal validity.

In implementation, baseline stability is demonstrated for the first leg of the MBL, and then the B phase is initiated for this leg of the design. All other legs remain in an extended baseline that continues until treatment of the first leg is completed. Then the experimental manipulation shifts to the second leg of the MBL. This leg had been in an extended baseline period, but now treatment is instated. Meanwhile, the remaining third leg of the MBL continues in baseline and

 ©Judith A. Gierut, 2015

remains there until treatment of the second leg is completed. Then the manipulation shifts again, with treatment delivered to the final leg. Thus, the AB sequence of extended baseline followed by treatment is applied sequentially for each leg of the MBL design.

With the MBL, a key assumption is that change will not occur in successive legs of the design until treatment is instated. With each baseline–treatment replication, it is further assumed that there will be a direct replication of effects to demonstrate causality between the delivery of treatment and change. Finally, because baselines are time-lagged, there is a further assumption that the effects of treatment are not due to extraneous variables. The logic is that, if extraneous factors were influencing learning, then they would apply uniformly across all legs of the design. In other words, "noise" associated with interference presumably impacts an entire system. In the study of acquisition, this levels claims that maturation is responsible for behavioral change in such studies. If change were due to maturation, then all legs of the MBL would be expected to mature accordingly. However, if the only behavior that advances is the one being treated, then maturation cannot be responsible for the gains. The distinguishing properties of the MBL are its demonstration of control through a time-lagged sequence of no change and the direct replication of experimental effects.

There are three common MBL applications, where generalization learning is explored across behavior, subject, or setting. As applied to language, the MBL across behavior manipulates legs of the design to induce changes in the phonological system. The MBL across subjects applies each leg of the design to a different child. The MBL across settings application varies the situation of treatment across legs of the design.

Learnability Project Experimental Protocols

The presentation of designs has emphasized learning during the experimental phase of treatment. However, of greater importance is the demonstration of generalization as the dependent variable. Generalization is monitored during treatment using a subset of the baseline items as the probe measure. Items that are probed are reserved specifically as a measure of transfer; they are never taught or presented during the teaching sessions. In the development of an experiment then, an investigator must detail two protocols, one that guides the delivery of treatment and another that determines how change is measured. These constitute the *treatment and generalization protocols* respectively. Typical treatment and generalization protocols for multiple baseline applications of the Learnability Project are outlined below.

Sample Treatment and Generalization Protocols of the Learnability Project

Pretreatment phonological analysis
- Administer entire PKP, OCP, CCP
- Establish interjudge transcription reliability

Multiple baselines as determined by experimental assignment
- Administer subset of PKP, OCP, CCP
    - Establish interjudge transcription reliability

Treatment phase 1: Train independent variable in imitation
- 75% accuracy x 2 consecutive sessions OR 7 sessions
- Treatment probes administered on VR2 schedule

 ©Judith A. Gierut, 2015

Establish procedural reliability of treatment administration

Phase shift generalization probe
- Administer entire PKP, OCP, CCP
- Establish interjudge transcription reliability

Treatment phase 2: Train independent variable in spontaneous production
- 90% accuracy x 3 consecutive sessions OR 12 sessions
- Treatment probes administered on VR2 schedule
- Establish procedural reliability of treatment administration

Immediate posttreatment generalization probe
- Administer entire PKP, OCP, CCP
- Establish interjudge transcription reliability

Two-week posttreatment generalization probe
- Administer entire PKP, OCP, CCP
- Establish interjudge transcription reliability

Two-month posttreatment generalization probe
- Administer entire PKP, OCP, CCP
- Establish interjudge transcription reliability

Details of a Sample Treatment Protocol

In the sample protocol shown above, there are two phases of instruction. These are delivered in 1-hr individualized sessions, three times weekly. Imitation is the first step, during which a child repeats a clinician's model of the treated items. The number of treated items varies depending on the experimental question. Items are arranged in blocks, so that a child produces each of the items as a discrete trial, before a second block of imitation trials begins. The accuracy of each trial is scored as a measure of trial-by-trial learning. On average, children produce about 100 trials per treatment session. During treatment, feedback is provided about the accuracy of sound production and corrective models are given. Imitation treatment continues until the child achieves 75% accuracy of production over 2 consecutive sessions, or completion of 7 total sessions, whichever occurs first. Treatment then advances to the second step of spontaneous production of the same treatment stimuli. The child produces each item independently, without a model. The procedures for blocking items, recording trial-by-trial accuracies, and delivering feedback remain the same as in imitation. The criterion for completion of spontaneous treatment is more stringent, set at 90% accuracy of spontaneous production over 3 consecutive sessions or until completion of 12 total sessions, whichever occurs first. The experimental phase of treatment does not typically exceed a maximum of 19 sessions (hours). Fidelity in administration of the protocol is established using a checklist procedure.

The specific methods of teaching vary by experimental question and included training production using single words, minimal pairs, and nonwords. Treatment of nonwords is a hallmark of our research, dating back to Gierut (1990). Nonwords are phonotactically permissible forms, balanced for canonical structure, vowel context, stress, and syntactic category. The segmental composition of nonwords is tailored to an individual child's

©Judith A. Gierut, 2015

phonological needs. Nonwords are introduced in children's stories to provide lexical support for learning. Nonwords correspond to character names or unusual objects or actions that take place in the stories. At the start of each treatment week, the nonword story is presented to the child prior to production practice and is then followed by the imitation (then spontaneous) phase of training on nonword production. Initially, we developed this paradigm as a means of controlling the treatment experience across children. Because nonwords are unique, none of the children have prior exposure to, or practice in saying the nonword sequences. Moreover, the same visual stimuli can be used for all children, varying only the phonological content as based on a child's phonology. Clinically, there is a precedent for the use of nonwords in treatment, dating to original models of treatment introduced by Van Riper (1963) in speech-language pathology. Recent advances in psycholinguistics and spoken word recognition lend further credence to treatment of nonwords. One hypothesis is that spoken word recognition takes place dually at lexical and sublexical levels. In lexical processing, details about the word as a whole are extracted, e.g., a word's frequency. In sublexical processing, details about phonological structure are the focus, most notably, the phonotactic probabilities of sounds and sound sequences. Because nonwords do not have lexical status, the emphasis of nonword treatment is at the sublexical level of processing. Moreover, because children with phonological disorders have difficulty learning the specifics of the target sound system, the use of nonwords may be especially beneficial. In fact, one hypothesis is that the locus of the problem for these children lies at the level of sublexical processing in word learning.

The treatment protocol thus specifies the details of day-to-day instruction, with decisions about the stimulus materials, number of trials and teaching steps, and methods of presentation, feedback and advancement. The data that emerge from the treatment protocol serve an important purpose in documenting that a child learned from the instruction.

Details of a Sample Generalization Protocol

In a generalization protocol, decisions must be made about how to measure change and to some extent, when to measure change, although this is largely dictated by the experimental design. The generalization probe must be sufficiently rich to cull information about change in treated and untreated properties of children's sound systems. Probes must also be ecologically valid and easy to administer and score. Probes should resemble the treatment in part, but must stand alone as a reflection of change; this controls for the potential of teaching to the test. Because probes are administered frequently, it is necessary that the learner's responses do not become stimulus or lexically bound. To guard against this, the probe should include multiple exemplars and contexts in sampling behavior. The Learnability Project typically relied on the Phonological Knowledge Probe (PKP), Onset Cluster Probe (OCP), and Coda Cluster Probe (CCP) as primary measures of phonological generalization. Multiple renditions of each relevant sound and cluster are elicited to prevent lexical binding, and multiple copies of the probes are used longitudinally to prevent stimulus binding.

The Learnability Project protocol necessitates that generalization probes are administered at key points in a single-subject experiment. During the A phase, probes are administered as the baseline measure. During the B phase, the same probes are again administered to monitor change as a function of treatment. Administration may involve the entire probe measure or a smaller subset of items. For example, the Learnability Project employed the PKP to measure the overall phonology at baseline, but only those items relevant to the experimental manipulation of the child's phonology were repeatedly probed for change as a function of treatment. Typically, sounds excluded from the child's phonemic inventory were monitored for change, with the number of probes administered being dictated by the experimental design.

 ©Judith A. Gierut, 2015

Probe administration may follow a fixed or variable schedule. A fixed schedule of probing might be implemented, for example, in sessions 3, 6, and 9, with probes set for every third session of treatment; whereas a variable probe schedule might be implemented in sessions 1, 2, and 6, where sampling averages every third session. The advantage of a variable schedule is that it is less predictable in time as a potential cue to testing. The protocol of the Learnability Project often used a variable probe schedule averaging 2 sessions. The probe is reserved exclusively for testing and feedback about accuracy of a learner's response is never provided.

From a design perspective, the most important probe administration occurs immediately at posttreatment because these data confirm the causal effects of treatment. Any probes that are administered after treatment is withdrawn bear only on the maintenance of treatment effects. If generalization takes place after treatment is withdrawn, these effects cannot be attributed to the treatment itself. The reason is that treatment as the independent variable is no longer operative, leaving open the possibility that some other factors may be contributing to continued learning. Nonetheless, continued probing for generalization is often used to plot trajectories of longitudinal change in documenting maintenance of treatment effects.

The Learnability Project typically collected probe data at baseline, phase shift from imitation to spontaneous production, immediately posttreatment, and at two weeks and two months after treatment was withdrawn. The PKP was conventionally administered at these points as a benchmark of system-wide generalization across the phonological system as a complement to session-by-session performance in treatment. In all, the generalization protocol provides the necessary evidence of change in treated and untreated sounds, within and across untreated words and contexts. This served as a primary dependent variable of single-subject design applications used by the Learnability Project.

References Cited

Gierut, J. A. (1990). Differential learning of phonological oppositions. *Journal of Speech and Hearing Research*, *33*, 540-549. PMID: 2232772

Van Riper, C. (1963). *Speech correction: Principles and methods*. Englewood Cliffs, NJ: Prentice-Hall.

For more information, visit the Gierut / Learnability Project at IUScholarWorks.

 ©Judith A. Gierut, 2015