

Generalizability Theory Applications and Extensions

Patrick E. McKnight (George Mason University; pmcknigh@gmu.edu)

Lee Sechrest (University of Arizona; sechrest@email.arizona.edu)

Overall Abstract

Generalizability theory (g-theory) is a powerful data analytic approach that can be applied to many scientific questions. The basis of generalizability theory is to estimate sources of variability to determine which sources affect outcomes most and where to focus future research efforts. Generalizability theory originally was intended to address questions in educational research, but the developers stressed its utility beyond this domain although there remain few examples of its application outside education. The lack of popularity may be due to the lack of mainstream statistical software, inadequate information about the procedure, or few specific examples of generalizability theory applications. Readily available software is not necessary but can be helpful to initiate new users whereas clear information and specific examples are essential to entice researchers to consider generalizability theory as a useful tool. The purpose of this symposium is to introduce g-theory – its use and purpose - and to provide specific examples of generalizability theory applied to experimental research, naturalistic studies, instrument development, and survey research in the areas of cognitive psychology, education, and medicine. Our objectives are to inform a diverse audience about the utility of g-theory for addressing a variety of common research questions and to demonstrate the application of a simple g-theory analytic tool (written in the R programming language) across different study designs and domains to address these questions. The demonstrations are supplemented by a primer session on conducting the analytic procedure and interpreting the results. Taken together, we aim to increase the awareness of our colleagues about this important data analytic approach and increase its application beyond educational research to a wider array of important research questions.

A primer and critical examination of generalizability theory

Patrick E. McKnight (George Mason University; pmcknigh@gmu.edu)

Lee Sechrest (University of Arizona; sechrest@email.arizona.edu)

Abstract

It is difficult to interpret generalizability theory results without a reasonably deep understanding of the procedure. That deep understanding may come from repeated exposure to results and critical feedback from reviewers; a process that takes years to unfold. Another way to develop a deep understanding can take a more direct path and provide a clearer understanding of these results – applying generalizability theory to manufactured data. The use of manufactured data with specific properties has a long-standing history in statistics and its utility goes beyond the development of new procedures. By using manufactured data, analysts can come to a deeper understanding of statistical procedures without the toil of trial and error. Our intent is to demonstrate the utility of generalizability theory while also clearly articulating how sources of variability can be interpreted as indicators of mean, variance, and covariance differences. Six small manufactured datasets with two (persons and variables) or three facets (persons, variables, and occasions) clearly demonstrate that generalizability results are reflections of mean differences and can be interpreted with ease once the data analyst realizes this point. Those mean differences are what we refer to as within facet variance but to the typical researcher they appear to be nothing more than mean differences within factors in an ANOVA table. By carefully reviewing the structure of these six data examples, we step through the output to show how g-theory captures the variance within facets and helps us better understand the nature of our effects.

Generalizability theory in laboratory interruptions research: Estimating variance to improve future research

David Kidd (George Mason University; dkidd3@gmu.edu)

David Cades (George Mason University; dcades@gmu.edu)

Patrick E. McKnight (George Mason University; pmcknigh@gmu.edu)

Abstract

Laboratory settings offer the researcher a distinct advantage of experimental control. The purpose of experimental control is to minimize the influence of extraneous variables and maximize the ability to detect the effects. If these extraneous variables are not controlled, the effects of a manipulation become increasingly obscured by noise and error. But how does the experimentalist know if he or she has sufficiently controlled for these confounds? Most researchers rely on standard methods, however, generalizability theory may offer researchers a way of explicitly quantifying different sources of variation.

Generalizability theory, while infrequently used by experimentalists, may shed light on these sources of variability and may make future research more profitable. The purpose of this presentation is to demonstrate the usefulness of generalizability theory in laboratory work on interruptions in human tasks.

Interruptions – a disruption of a focal task - are known to negatively affect task performance by delaying resumption and increasing the likelihood of task errors. The current study investigated the effects of brief interruptions on task performance. Eighteen participants were interrupted by interruptions ranging from 250 ms to 5000 ms while completing a primary task. It was expected that task resumption would be faster for shorter interruptions, however, initial statistical analyses were inconsistent with theoretical predictions. We used generalizability theory to identify the sources of variability in the effect to help inform future research decisions. The results of the generalizability analysis indicated that mean differences were mostly due to between subject variability and not the experimental manipulation. These results suggest that improved sampling procedures and more careful manipulation of interruptions ought to maximize the expected effect.

Where is the real-world variance? A generalizability theory approach to understanding interruptions in naturalistic environments

David Cades (George Mason University; dcades@gmu.edu)

David Kidd (George Mason University; dkidd3@gmu.edu)

Patrick E. McKnight (George Mason University; pmcknigh@gmu.edu)

Abstract

Social science exists on a methodological continuum ranging from the basic, controlled laboratory setting to less controlled and more applied naturalistic setting. The primary differences between the laboratory and naturalistic studies are tradeoffs between strength of inference and ecological validity. A question that both laboratory and naturalistic researchers constantly face is how to balance the strengths and weaknesses of both approaches. Generalizability theory may be a tool that can help span this gap by adding statistical power and control to directly applied work. The following presentation addresses the applied side of this continuum with data from a naturalistic experiment on interruptions.

As laboratory studies have shown, interruptions can be deleterious to our ability to complete tasks, having negative effects on both time and accuracy. Given these basic research findings, how can we understand the role interruptions play in real-world tasks? One method is to observe people in their natural environments and assess their performance in dealing with interruptions. Ten hours of data from administrative office workers were recorded and analyzed with a focus on performance with interruptions. We identified several relevant variables (interruption length, environmental changes, degree of task importance) and analyzed the results with generalizability theory; the purpose was to provide evidence about the specific variables that most greatly influenced performance. Applying generalizability theory to this data enabled us to understand the effects at a level usually only reserved for highly controlled laboratory studies. Additionally, changes to the environment proved to contribute a large proportion of the variance in performance with interruptions. These results lead to a more focused naturalistic data collection strategy along with the design of tools and training programs that mitigate the negative effects of interruptions. This application of generalizability theory may illustrate how to use laboratory analyses in noisy, uncontrolled naturalistic environments.

Assessing the dependability of judges in an observational task: The role of generalizability theory in instrument development

Katherine M. McKnight (Pearson Education; kathy.mcknight@pearson.com)

Patrick E. McKnight (George Mason University; pmcknigh@gmu.edu)

Claude Goldenberg (California State University – Long Beach; cgolden@csulb.edu)

Rhoda Coleman (California State University – Long Beach; rhoco@earthlink.com)

Anastasia Amabisca (California State University – Long Beach; aamabisc@csulb.edu)

Abstract

The purpose of this talk is to illustrate the use of g-theory to address a common psychometric issue in instrument development: raters' consistency in their use of a given instrument. In this study, we developed and tested an observation protocol designed to rate teacher classroom behaviors related to instruction of English Language Learners (ELLs). The DIELOP ("Differentiating Instruction for English Language Learners Observation Protocol") is an 11-item observation tool used to judge the quality of specific teaching elements during a teacher-directed lesson. There were two types of ratings for each item: *quality* ratings, ranging on a scale of 0 ("not observed") to 5 ("very effective"), and *quantity* ratings which yield a frequency count based on the observed number of a defined set of behaviors (ranging from 3 to 10) associated with each item. Six items reflect standard teaching functions associated with effective directed lessons (e.g., a discernable objective) and 5 items are geared specifically toward instruction of ELLs (e.g., adapting instruction to English proficiency levels). The purpose of the observation protocol is to evaluate whether a specific intervention is effective for improving instruction for ELLs.

To test the DIELOP, we observed 10 teachers using three independent raters and analyzed the results using a g-study approach. The analysis assessed whether the raters were interchangeable, i.e. whether they used the instrument in the same way for both quality and quantity ratings. We tested 3 sources of variability: items, teachers and raters. Results indicated that raters contributed negligibly (less than 2%) to overall variance in quality and quantity ratings, and teachers contributed the most (32% for quality, 14% for quantity ratings). The greatest source of variance was the 3-way interaction (item x rater x teacher). We discuss the implications for further development and use of the DIELOP.

Generalizability theory in survey research: A critical analysis of items and respondents

Patrick E. McKnight (George Mason University; pmcknigh@gmu.edu)

Lee Sechrest (University of Arizona; sechrest@email.arizona.edu)

Abstract

Typical survey research results in tabular data of item responses and probability estimates; these results can be greatly enhanced by the application of generalizability theory. Results such as the dependability of respondents across individual items, item domains, and response categories constitute just a few of the questions that may be relevant to address. Dependability across items provides evidence about the utility of multiple items whereas dependability of items across domains or persons across domains informs us about the utility of clustering survey items. These results, while not traditionally conducted in survey research provide unique but important insights into the nature of survey response and the general utility of those responses.

We analyzed data from a recent national survey of medical decision making – the National Survey on Medical Decision or NSMD; a survey conducted for the purposes of better understanding the individual medical decision process in the United States health care system. The data included three separate item modules (screening, medications, and surgery) administered partly across roughly 3000 respondents to estimate the sources of variability in a three-facet model – respondents, items, and modules. Results indicated that respondents tend to provide generalizable results across modules but not across items suggesting that people tend to use the same judgements about different medical decisions regardless of the medical domain (e.g., screening or surgery). As indicated, the items did not generalize well - indicating that the survey questions offered unique sources of variability. Furthermore, respondents differed substantially indicating that the survey sampling procedure offered sufficient variability. These results offer a glimpse at the power and utility of generalizability theory in survey research and may help in both current analyses of survey results as well as future survey designs.

Generalizability Theory Analysis for Psycholinguistic Applications

Aurelio Jose Figueredo (University of Arizona; ajf@email.arizona.edu)

Sally Olderbak (University of Arizona; sallyo@email.arizona.edu)

Abstract

We propose that continuing controversies over the use of quasi-F-ratios in psycholinguistic research might be circumvented, if not resolved, by judicious application of Generalizability Theory (GT) analyses. We argue that GT is a logical extension of the basic rationale behind repeated measures Analysis of Variance (ANOVA) and the variance components model upon which GT is ultimately based and upon which the entire logic of the F-ratio (quasi or otherwise) rests. GT is especially useful in psycholinguistics research because it affords one the opportunity to assess generalizability across multiple dimensions within the same model, such as individual subjects as well as varying conditions of prime and target words. We will provide an illustrative example of GT based on Forster's (2007) replication of Davis and Lupker's (2006) study in which they tested the effects of frequency discrepancies in target and prime words across individual subjects under varying combinations of frequency and prime.

With advance apologies to the statistically more sophisticated audiences, we approach this matter didactically by briefly reviewing a few of the basic concepts of Analysis of Variance (ANOVA) so as to develop a common symbolic lexicon with which to build a bridge to Generalizability Theory (GT) Analysis. By building on the well-known, we hope to shed some light upon the less familiar and render it more accessible to a wider audience.