

## Insights in Language Testing:

# An Interview with Robert C. Gardner

by Parrill L. Stribling



Robert Gardner is a Professor Emeritus at the University of Western Ontario. He obtained a Ph.D. in psychology from McGill University in 1960 and started teaching at the Univ. of Western Ontario the following year. A leading authority on attitudes towards second language acquisition, Gardner has also been crunching numbers for many years. The structural equation modeling (SEM) approach in his 1972 text *Attitude and motivation in second language learning* was absolutely stunning. Gardner applied a new statistical procedure to language learning attitudes and explained his methodology in a clear, straightforward manner. His recent *Psychological statistics using SPSS for Windows*, this newsletter's featured book review, is an outstanding work on statistical analysis and methodological procedures. This interview was conducted by email in the winter of 2003.

## Part I - General Questions

### ***What changes have you seen in the testing field since starting your career?***

When I started there was a distinction between ability and affective tests, and that distinction still exists. Generally, ability tests (which include measures of intelligence and achievement) are performance measures. Individuals get answers correct or not. It is easy enough to fake bad, but not so easy to fake good. More times than not, affective tests (which measure personality, attitude, mood, motivation, etc.) consists of verbal report which can be answered in many ways. That is, individuals can give socially desirable responses; they can fake good or bad; or they may acquiesce, etc. Since starting my career, techniques have been developed to attempt to identify response biases and acquiescence, and to counteract or identify social desirability responding and the like. However, many of the same problems remain.

More recently, there has been the development of item response theory (IRT) and advances in computer adaptive testing, and even more biologically based assessment measures (e.g., FMRI, etc.). Time will tell how successful these will be in improving our assessment of important individual characteristics that help us predict behaviours. There have also been many advances in data analysis, ranging from item analysis techniques to factor analysis and structural equation modelling.

***Do any recent trends in testing concern you a lot?***

As far as the concern for measurement adequacy is concerned, most of the developments have been promising, and there are more opportunities in the future. Perhaps its my age, but suddenly things seem to be changing more rapidly now than ever before. Or have we always said that?

*“the impromptu development of tests based on data analytic procedures . . . are not positive trends in testing, but rather events stemming from the easy accessibility to powerful data analytic procedures and a lack of formal training in statistics and test construction.”*

One concern is the impromptu development of tests based on data analytic procedures such as factor analyses of items and the identification of "tests" based on exploratory factor solutions seem to be increasingly frequent. These are not tests based on theory or a careful selection of representative items and the like, but rather on the happenstance of a factor structure or contributions of a regression equation. In my view, these are not positive trends in testing, but rather events stemming from the easy accessibility to powerful data analytic procedures and a lack of formal training in statistics and test construction. Happily, tests developed by such post hoc procedures do not last long, but they do introduce unnecessary distractions in the quest for understanding individual differences. Perhaps that's a small price to pay for the availability of so many ways of sorting data.

***Do psychometric statistics differ from the way statistics are used or interpreted in other disciplines?***

I don't think psychological researchers use or interpret statistics any differently than those from other disciplines, but it is possible that they are more familiar with statistical procedures and use them more extensively than researchers do in other disciplines.

***Part II - Questions about Psychological statistics using  
SPSS for Windows***

***Should a student be familiar with all of the analysis measures in your book Psychological statistics using SPSS for Windows to obtain a post graduate degree which includes a quantitative study?***

I began writing this book in 1993 when asked to develop a course for an undergraduate course in data analysis. At the University of Western Ontario, psychology undergraduates take one full course in statistics, and those wishing to honor in psychology take a second course dealing with computer applications in psychology. At that time there wasn't a text that combined the rationale of statistics with computer applications, so I began with two chapters, one on factor analysis and the other on multivariate analysis of variance, because these topics were not covered by most undergraduate texts. The book grew from there covering the topics that seemed the pertinent to student needs, and in 1998 I finally submitted a complete draft to the publisher. Though the initial statistics course at UWO was optional, but by 1996 it was made compulsory largely because students felt that it was necessary to be able to do their undergraduate theses.

The short answer to your question is that graduate students interested in research careers would need to know the material in this book as a bare minimum. If I were writing a statistics book for language researchers, I would also add some chapters on topics such as complex (i.e., 3 and 4 factor) analyses of variance, analysis of covariance, reliability and validity assessment, and the fundamentals of structural equation modelling. And of course, I would change the examples to focus on language related issues.

*“. . . a purely cookbook text on data analysis is of limited value . . . it is important for students to learn the rationale, assumptions, and limitations of various analytic procedures.”*

Although I'm a professor emeritus, I still teach the graduate course in research design and though much of my course deals with analytic procedures, a more important focus is the mathematical fundamentals and the interrelationships among the

various procedures. My feeling is that a purely cookbook text on data analysis is of limited value. Students should understand what it is they are doing and why. Simply following examples can often be misleading, so it is important for students to learn the rationale, assumptions, and limitations of various analytic procedures.

***One user-friendly aspect of Psychological statistics using SPSS for Windows is that each chapter has a short table of contents and a bit of history for each analysis method discussed. Why did you choose to organize it like that?***

When I read articles and books, I like to have an overview of what they are about before starting. Though journal articles don't have tables of contents, they do have abstracts and subheadings, so I will read these carefully before reading the article itself. I find that it helps me to conceptualize articles more clearly. When I wrote the book, I thought it would help if I simply preceded the article with the section headings. When I used the book as a text, I told the students that they should study the table of contents before they read the chapter to help to give it more structure.

The reason I added the bit of history sections is because I have noted over the years that students somehow just think statistics has always been here and that everything is cut and dried. I didn't want to put in too much detail but I did think it was important for students to see when the various procedures were introduced (you will note that many of them are less than 100 years old), and the context in which they evolved. The details are, however, very interesting, and I would recommend that readers investigate Cowles' *Statistics in psychology: A historical perspective*. It is well written, interesting and easy to read. When I used my book and its predecessors in my undergraduate class, I was frequently asked whether the historical material would be on the exam. There was always a look of consternation on the faces of the students when I said "yes", as if somehow this wasn't right for a statistics class.

***What type of reader do you see benefiting most from Psychological statistics using SPSS for Windows?***

I think you can read the book at many levels. I have received positive comments from students who had not used computers for data analysis and who had just a basic course in statistics, and from colleagues who are very knowledgeable about statistics. Actually, I believe it could be used profitably as a cookbook by individuals with very little statistical knowledge as well as by students who want to know the fundamentals of the various procedures, and by skilled researchers who may not have done one of the analyses types mentioned in the book recently.

### **Part III - Analytical Methodology Questions**

***Why are Standardized Regression Coefficients and Structure Coefficients so difficult to interpret (pp. 209 - 213, pp 221 - 222)? Is the usage of part and partial correlation coefficients in a regression equation an example of fuzzy psychometric usage? Why do you caution interpreting part and partial correlation coefficients?***

I have grouped these three questions together because they are highly interrelated. The best answer I can give to them is in Chapter 9 of the text. The short answer is that multiple correlation involves the correlation between one variable, the criterion, and a weighted aggregate of a set of variables (the predictors) where the weights are determined to make the correlation as high (and positive) as possible.

The unstandardized regression coefficients are simply the weights applied to the raw scores, while the standardized regression coefficients are the weights applied to the scores in standard score form. Furthermore, the standardized regression coefficient is simply the unstandardized regression coefficient multiplied by the ratio of the standard deviation of the variable in question to the standard deviation of the criterion (see page 211 of my book), and you can prove this to yourself by performing the arithmetic on any multiple regression analysis you have run.

Both standardized and unstandardized regression coefficients can vary beyond the ranges of plus or minus one depending on the relationships among the variables (particularly the "predictors"), and the standard deviations of the variables. Although it is true that in most instances the standardized regression coefficients tend to vary between minus one and plus one, this isn't necessarily the case. As a consequence, interpretation of either the standardized or the unstandardized regression coefficients is hazardous because you don't really know what is large or small. Also, as I discuss in my book, a larger standardized regression may not even be significant whereas a smaller one is, so relative comparisons are also hazardous.

Unfortunately, it is extremely common in many fields of research to interpret the regression coefficients as if they mean something. I have recommended in the book that if you really want to interpret the regression coefficient, you instead interpret either the part or the partial correlation, and I show in my book how you can compute the part correlation using the  $t$ -value for the regression coefficient, the squared multiple correlation and the degrees of freedom for the error term (see pp. 212, 221). There is also a formula for the partial correlation, but this is a bit more complex so I didn't include it in my book. In my graduate class, I encourage students not to interpret the regression coefficients, but then capitulate and say that if they must (because I realize

*SHIKEN: The JALT Testing & Evaluation SIG Newsletter, 8 (1) March 2004. (pp. 11 – 21).*

someone will pressure them to do so), would they please interpret either the corresponding part or partial correlations instead. The reason for this is that to present a part or partial correlation it is necessary to say that the value in question is the correlation between the criterion and the particular predictor once the variation due to the other predictors in the equation is removed from the predictor (part correlation) and the criterion (partial correlation). The other thing, of course, is that part and partial correlations can vary only from -1 to +1. As an aside, I might note that the tests of significance of the regression coefficient, the part, and the partial correlation are all equivalent. The formulae often given to determine significance look different, but they can be shown to be algebraically equivalent. That is, if the regression coefficient is significant at say the .032 level, the other two will also be significant at the .032 level.

***What about using regression coefficients in Structural Equation Modeling? Isn't this what you did in your book *Social Psychology and Second Language Learning: The Role of Attitudes and Motivation?****

That's very perceptive of you and a very interesting question. Structural equation modeling represents a test of a model. The coefficients apply only to the variables in the model, thus when you present it, you are basically saying that if this is an appropriate model given these variables, then this is the nature of the regressions of the various variables on other variables. The model is internally consistent. When interpreting the model, attention is directed toward significant paths (generally given in standardized form and considered significant when the critical ratio is greater than plus or minus 2), but only the sign and the significance is considered because of what was said above. Even in this situation the standardized coefficients can exceed an absolute value of 1, though often this can be overcome by doing what is called a "completely standardized solution". This doesn't really eliminate the problem, but the point is that the magnitudes of the coefficients are not the focus, only the nature and signs of the paths.

***Are unstandardized coefficients a stronger indicator of a variable's strength in a regression equation than standardized coefficients?***

Neither is a stronger indicator than the other. They are equivalent. As I have said one is simply the other multiplied by the ratio of two standard deviations. The unstandardized regression represents the weight in raw score form, and corresponds to the slope of the criterion against the predictor when all the other predictors in the equation have been residualized (i.e., have been made independent of one another,



*SHIKEN: The JALT Testing & Evaluation SIG Newsletter, 8 (1) March 2004. (pp. 11 – 21).*

regardless of their intercorrelations). These slopes refer to variables that have different standard deviations (and means, for that matter). The standardized regression coefficient is the weight in standard score form. It is the slope of the standardized criterion against the standardized predictor when all the other predictors in the equation have been residualized (i.e., have been made independent of one another, regardless of their intercorrelations). These slopes refer to variables that have standard deviations of 1, (and means of 0).

***I have found the SPSS regression output confusing to interpret. The table you give as an example (on p. 220) is what I spent a long time on my own research trying to understand. Do you think SPSS labeling of Beta for Standardized Coefficients and B (which really are the Beta coefficients) is confusing? Are Beta coefficients unstandardized?***

I don't think it's the SPSS regression output that is so confusing to interpret, I think it's the whole issue of what is multiple regression and what does it really tell us. Often, people say that they are using multiple regression to find the best predictors, but multiple regression does not provide this information. Multiple regression identifies those variables that add to prediction over other variables in the equation, and that is a different matter altogether. A variable can be a very poor predictor (i.e., not correlate significantly with the criterion), but can add significantly to the prediction achieved by other variables in the equation. And this is what leads to all the interpretation difficulties discussed above. As for the labelling in SPSS, *Beta* is the standardized regression coefficient, and *B* is the unstandardized regression coefficient. The only confusion is that in most textbooks, the unstandardized regression coefficient is usually identified with a lower case *b*.

*“Often, people say that they are using multiple regression to find the best predictors, but multiple regression does not provide this information.”*

***I have never obtained varying results between a Forward or Backward selection. I noticed that the analyses would reject a variable and the only way to obtain the part and partial coefficients for all the variables in a regression was to use Enter because the analysis was forced to compute the part and partial coefficients of all the variables. Why would someone choose to use either a Forward or Backward selection if there is no difference in the outcome?***

*SHIKEN: The JALT Testing & Evaluation SIG Newsletter, 8 (1) March 2004. (pp. 11 – 21).*

What you say is true, especially if there are a relatively few number of predictors and/or a fairly simple structure of relationships. With a large number of predictors and a very complex set of predictors, you can identify some different variables as contributors depending on whether you use Forward Inclusion or Backward Elimination. But the question one might ask is so what? With Forward Inclusion you let the computer decide which variable correlates the highest, then which has the highest part correlation once the first predictor has been partialled out, then which when the first two predictors are partialled out, etc... In the end, you have an equation for which not all of the predictors may have significant regression coefficients, but they did along the way. With Backward Elimination, you enter all variables then eliminate the one that has the smallest (in an absolute sense) and non-significant  $t$ -value for the regression coefficient, recompute the regression equation with that variable eliminated, and then eliminate the variable with the lowest absolute  $t$ -value, etc... With this approach, you will have variables that all have significant regression coefficients on the final step, though if you use the SPSS Backward default option, the  $p$  value to retain is less than .10. In my book, I recommend against using any of the indirect solutions, and this is true of most people who write books on the use of multiple regression. The problem with any of these approaches is that they capitalize on all the chance variation in the sample of data and most likely will not replicate on another sample of data using the same variables.

***The rotation issue for factor analysis is explained (on pp. 244 - 245), but why do some people such as Kline (1994) state that a factor analysis is incomplete without a rotation, would you agree? And, if so, does this imply that research involving a factor analysis could be considered invalid, if a rotation was not performed?***

*“The whole rationale in rotation is to try to distribute the variance more equally among the factors . . . thus making for a more parsimonious or psychologically meaningful interpretation.”*

It really depends on your purpose. It is common to use rotation algorithms because these often offer a more parsimonious or interpretable solution. Many factor analytic solutions extract factors in decreasing order of the amount of variance in the matrix accounted for by each factor. Thus, if you were interested in these relative amounts of variance, you would probably opt for an unrotated solution. These are more complex to interpret, however. The whole rationale in rotation is to try to distribute the variance more equally among the factors (if that is consistent with the



*SHIKEN: The JALT Testing & Evaluation SIG Newsletter, 8 (1) March 2004. (pp. 11 – 21).*

relationships among the variables), thus making for a more parsimonious or psychologically meaningful interpretation. Both the rotated and the unrotated factor solutions will however, reproduce the same correlations among the variables, so there is not really any difference in the relationships described; only in the language used to describe the relationships. If one wants a simple description of the relationships, this is most often provided by a rotated solution (particularly an orthogonal one).

***You point out that replication will prove if the sample size effects factor analysis (p. 243), but some people state five cases per variable and others will go as low as two. What would be a good rule of thumb for the novice?***

On page 243 of my book, I cite references that suggest that the sample size should be anything from 2 to 20 times greater than the number of variables, but then briefly describe research that points out that a more important feature is the underlying structure. If it is simple and well defined, sample sizes can be smaller than if it is complex and/or ill defined. The real issue, in my opinion, is whether or not the results are stable on replication. In my own research, I have often done factor analyses with about 20 variables, and I have generally found that the factor structures as reflected in the interpretations of the factors are fairly consistent when the sample size is about 100. Thus, you might say that I support the 5 to 1 ratio. However, if I had appreciably more variables, or if I expected a fairly complex factor structure, I would certainly increase the sample size, and possibly to a ratio greater than 5 to 1.

***Is it a good idea to run a principal components factor and determine the number of acceptable factors based on the scree plot? What alternative method are there for determining the number of acceptable factors?***

The default in SPSS Factor is the eigenvalue 1 criterion when performing a principal components or principal axis analysis. That is, all factors with an eigenvalue greater than 1 are retained. Often, particularly with a large number of variables, this will result in too many factors, since the eigenvalue is simply the sum of the squared factor loadings on each principal component. Because of this, the scree test is often preferred, because it will often show that the scree develops above an eigenvalue of 1. Some people claim, however, that this test is unreliable, in that different researchers identify the scree differently sometimes. Another possibility is to investigate the residual correlation matrix. If these values are all close to 0, or if they show a relatively normal distribution around 0, one can conclude that all the meaningful variance has

*SHIKEN: The JALT Testing & Evaluation SIG Newsletter, 8 (1) March 2004. (pp. 11 – 21).*

been extracted. There are a number of other approaches, however, that have been proposed, but generally they are computationally labourious or complex, so the scree and the eigenvalue 1 criterion are the most commonly used. One can get into quite an argument here, but generally I find that there isn't that much difference between the various criteria if they are followed with an eye on the nature of the variables making up the analysis. And, in the long run, the important point is will the solutions be comparable if the study is replicated. In the end, that is the most meaningful criterion.

### ***Can residuals be tested for Cronbach's Alpha reliability?***

This would depend on what you mean by residuals. The Cronbach *alpha* is a formula that refers to consistency in terms of a series of items. If you have  $k$  items, the Cronbach reliability coefficient is equal to  $(k \text{ over } (k - 1) \text{ times (the variance of the total scores minus the sum of the item variances ) divided by the variance of the total scores. This is difficult to express without an equation, but I hope it is clear. The point is that if you had calculated a total score by aggregating a set of residuals, then yes, you could compute a Cronbach } alpha$  by using this formula with the residuals. It would probably be the case that the coefficient would be low, largely because residuals tend to be unrelated to each other, but that is another matter.

## **Conclusion**

### ***What are you planning to write next?***

I've just submitted a book entitled *Analysis of variance with a continuous independent variable: Model I, the unique approach* to a publisher. It will be some time before I know whether it will be published. Personally, I think it is a much needed book for a computational area that is currently quite complex.

I do have hopes of writing one more book on the area of language attitudes and motivation, and am currently working on an idea. A colleague of mine from Spain and I have used our Attitude/Motivation Test Battery with of students in Spain learning English as a second language and obtained results that are very similar to those we have obtained over the years with Canadians learning either French or English. This has suggested to me that the oft-made comment that many of our results may be specific to Canada may reflect more variation in the nature of the items that researchers have used rather than the phenomenon itself. I am currently working on the idea of trying to obtain data sets from a number of countries. If successful, there might well be a useful book on the international use of the Attitude/Motivation Battery. Time will tell.

## **Works Cited**

Cowles, M. (2000). *Statistics in psychology: A historical perspective*. London: Lawrence Erlbaum.

Gardner, R. C. (2001). *Psychological statistics using SPSS for Windows*. Upper Saddle, New Jersey, USA: Prentice Hall.

Gardner, R. C. & Lambert, W. E. (1972). *Attitude and motivation in second language learning*. Rowley, Massachusetts: Newbury House Publishers.

Kline, P. (1994). *An easy guide to factor analysis*. London: Routledge.

HTML: [http://jalt.org/test/gar\\_str.htm](http://jalt.org/test/gar_str.htm) / PDF: <http://jalt.org/test/PDF/Gardner.pdf>

Copyright (c) 2004, by Robert C. Gardner & Parrill L. Stribling. All rights reserved.