Research Applications Of Computerized Measurement Of Response Times In Psychological Testing

Ken Stange, Nipissing University, North Bay, Ontario, Canada

Ken Stange
Nipissing University, 100 College Drive, North Bay, Ontario, P1B 8L7, Canada
Phone: (705) 474-3461 Extension 4524
Email: ken@stange.com

Abstract

One particularly fruitful application of computer technology in psychology is in the research area of computerized testing and measurement of response latencies (a variable often difficult to measure objectively and accurately by any other means). This brief report reviews some research applications the author has made of computerized recording of response times in the areas of personality testing, criterion-referenced testing, and empirical aesthetics.

Introduction

In psychological testing one of the more difficult dependent variables to measure accurately and unobtrusively by conventional methods is response time. The length of time it takes for a person to respond to something has been considered significant from the very beginnings of psychology as a science, with reaction time studies among the first truly objective investigations in psychology (Sternberg, 1969). Initially reaction time was believed to reflect actual neural transmission time to and from the central nervous system. Subsequent technological developments, which made possible direct monitoring of the time it takes for an action potential to travel along a neuron, have laid this misconception to rest. Most of the time involved in responding to a stimulus, even a simple stimulus requiring a single, simple response, is spent in the processing going on in the brain between arrival of information along the afferent neurons and the sending of the output along the efferent neurons. It is probably fair to say that afferent/efferent neural transmission time is a quite small constant added to the

total response time to anything, and that what makes response time interesting is the amount of time our brains take to process incoming information before instructing our fingers or tongues to do something.

An example of this in the differences found for reaction times to different kinds of judgment disjunctive reaction times. The number of choices in a disjunctive reaction time study increases the RT, as does the depth of processing required. Even more obviously, time to respond has long been correlated with difficulty on knowledge-based or aptitude testing, as indicated by the widespread use of time-limited tests. That even general intelligence has been widely associated with response times is reflected in the use of the terms "quick" or "slow" in describing a person's intellectual abilities.

Another aspect of a person, one related to personality rather than intellect is based on Freud's observation that response latencies to word associations can be an index of their emotional 'weight'. This was considered so important, that by 1887 Cattell had proposed a formal methodology for recording responses and response times to word association tests. It still remains important. Recent research in personality testing has linked response latencies to 'faking' (Holden & Hibbs, 1995).

Yet another area where response times have been shown to be of a useful objective measure of subjective processes is empirical aesthetics. One way of measuring a person's response to a piece of music or a work of visual art is by measuring the time voluntarily spent attending to it. Research in empirical aesthetics and evaluation judgments often has used subject-controlled viewing or listening time as an indicator of 'interestingness' or 'pleasingness' (Crozier, 1974).

Unfortunately, to accurately measure how long a subject takes to respond to something, be it a question on a test or the presentation of an image, is often difficult and cumbersome to do by conventional means—and it is especially difficult to do without the experimenter's obtrusiveness becoming a problematic extraneous variable. Experiments have been done where a movement detector on a subject's throat is connected to a timer to measure verbal response times, but it would be naïve to think that such apparatus isn't a confound. And how is an experimenter to measure response times to individual questions on a written test of skill or personality? Certainly looking over the test subject's shoulder, stopwatch in hand, is not ideal.

Computerized measurement of response latencies, on the other hand, is unobtrusive, entirely objective, and accurate—and doesn't even require the physical presence of data collectors. (There are accuracy problems when very precise measurements are made using the computer clock as addressed by Myors in 1998, but these are unimportant in most applications that are not dependent on millisecond precision.) Furthermore, the programming of experiments to present stimuli and record response times is relatively easy with multi-media, user-friendly languages such as Visual Basic. The algorithm for measuring response times is simple: read the computer clock when the computer presents the stimulus, read the clock when computer detects a response, and have the computer subtract the former from the latter and store this information.

What follows are three examples, from personal experience, of computerized response timings in the areas of personality testing, criterion reference testing, and empirical aesthetics—and some ideas for future research.

Example Of An Application In Personality Testing

According to conventional psychoanalytic wisdom (and conventional police interrogation wisdom), long or very short response times are an indicator of emotional 'weight'. If the psychiatrist (or interrogating cop) says "knife" and the person responds very quickly or very slowly, it is usually interpreted as meaning that a knife means something special to that individual. Extrapolating from this, it seems reasonable to expect that the response latencies to questions on a personality inventory would to some extent reflect the personality of the person taking the test. For example, since extroversion is often associated with spontaneity, one might expect that response times for an extrovert taking a personality test would be shorter than for someone who tests as more introverted.

To test this hypothesis, the Eysenck Personality Inventory was computerized. The program presented each question and timed responses. It was possible for one person to administer this test to several hundred people in just a few days—with the computer collecting all the data in computer file formatted and ready for statistical analysis in SPSS. Conducting this study without a computer, aside from the virtual impossibility of unobtrusive and accurate response timing, would have involved thousands of hours.

A potential confound in this sort of experiment is reading time: a subject could take longer to respond simply because of reading speed. It is hard to imagine a way of controlling this extraneous variable in a paper and pencil test. However it

was easy in the computerized version: the questions were programmed to appear on the monitor one word at a time at a rate with which any literate person could keep up.

As it turned out, the hypothesis that the two personality traits measured by the EPI (introversion/extroversion and neuroticism/stability) would relate in some way to response times to the different questions was *not* supported at a significant level. However, the fundamental assumption of emotional weight relating to response times *was* supported by an interesting finding regarding one of the questions. The original EPI has a question (weighted on the introversion/extroversion scale) about whether one would prefer reading a book to going to a "gay" party. Because the word "gay" has acquired a different meaning since the original test was devised, two versions of the EPI were given, one of which substituted the synonym "lively" for "gay". A highly significant difference in response times was found between these two versions of the question—with, of course, the "gay" version times being much longer. (Saltstone, Saari, Stange, & Walsh, 1989).

This whole area of the relationship of response times to personality characteristics and emotional "weight" is certainly worthy of further study. And it has become possible, even easy, because of micro-computer technology.

Example Of An Application In Criterion-Referenced Testing

It has been casually observed by many an instructor that those students who finish an exam early and those students who stay to the 'bitter end' tend to be at the extremes of the performance distribution—either doing extremely well or extremely poorly. The author has developed a commercial software package ("Scrutiny") that flags improbable similarity in selection of wrong answers among test-takers—a possible indicator of cheating by copying.

This application of computer technology is peripheral to the current discussion of response times, but nonetheless is worthy of mention as yet another example of the value of computers in opening up new areas of research. Although the program's primary function is flagging probable copying in a formal examination situation, it also does statistical item analysis. Again, the ease with which this can be done and the time-saving nature of using computers to collect and organize data ready for analysis is noteworthy.

The current program analyzes data files from scanned answer sheets, but a version in development will include the option of computerized administration of tests

and will record response times to each question. This will mean that every test or exam given will supply research data on the relationship of such variables as distracter strength and item difficulty to the time taken to answer a question. The insights into testing that such a wealth of data will offer is yet another example of the gift to science that are microcomputers.

Example Of An Application In Empirical Aesthetics

Judgments of the aesthetic value of art works traditionally have been made by having subjects use rating scales for variables such as 'interestingness', 'pleasingness" and "complexity". Another measure of aesthetic value sometimes used is the length of time a person chooses to attend (view or listen) to a work of art.

The author has created software ("DrMiro") for research on the relationship of fundamental visual components (colour, shape, size, and number of units) of an image to judgments of aesthetic value. The program generates random abstract images according to settings made by either the experimenter or the subject. It saves the actual images for retest-reliability studies. It can request Likert Scale judgments from the subject on each of the images generated. The program also measures viewing time (or latency to making a judgment) for each image created or being evaluated.

Although statistical analysis of several studies using this software is still underway, there is clear evidence of some relationship between the primary evaluation variables, actual time spent viewing the images, and re-evaluation reliability coefficients. One finding of considerable importance is that of reliability measures, including those based on viewing times, which suggests that much of the research based on assumptions that dependent measures of aesthetic value are reliable over time—not just circumstantial.

Perhaps most notable about the studies done using this software, as well as many studies that use computers, is the plethora of data easily collected. The massive amount of data so easily collected in any computerized study can result in a failure to see the forest for the trees. Furthermore, given enough data, one can always find something of 'statistical significance' as long as one is willing to sit at a computer and pump information into SPSS long enough.

Conclusion

Given the importance of response times in so much psychological research and the ease with which this variable can be measured in computerized experimentation and testing, it is expected that this measure will become standard (and extremely useful) in future research.

References

Cattell, J.M. (1887). Experiments on the association of ideas. *Mind, 12,* 68-74

Crozier, J.B. (1974). Verbal and exploratory responses to sound sequences varying in uncertainty level. In Berlyne, D.E. (Ed.) *Studies In The New Experimental Aesthetics* (pp. 27-90). New York: John Wiley and Sons.

Holden, R.R. & Hibbs, N. (1995). Incremental validity of response latencies for detecting fakers on a personality test. *Journal of Research in Personality, 29,* 362-372

Myors, Brett (1998). A simple graphical technique for assessing timer accuracy of computer systems. *Behavior Research Methods, Instruments & Computers*, 30(3), 454-456

Sternberg, S. (1969). Memory-Scanning: Mental Processes Revealed by Reaction-time Experiments. *American Scientist, 57,* 421-57

Saltstone, R., Saari, M., Stange, K., & Walsh, S. (1989). The Gay Party Question of the Eysenck Personality Inventory: An Example of Effects of Item Alteration on Direction and Latency of Response. *Perceptual and Motor Skills*, *69*, 735-738

Notes on Software Mentioned

A demonstration version of the program *Scrutiny* is available for downloading at…
http://api.simplenet.com/

The *DrMiro* software for empirical aesthetics research can be downloaded for free at…
http://kenstange.com/drmiro/