Measuring Sequences of Human Errors:  
Analysis of Patterns in Time

Jaesoon An and Theodore Frick

Abstract

Analysis of patterns in time (APT) is a quantitative method for measuring and analyzing qualitative patterns that are temporal sequences. In APT temporal relations are measured directly, in contrast to independent measurement of variables which are then related statistically. An empirical study illustrates the utility of such an approach, in which human errors in using modern software are predicted temporally from antecedent conditions. Implications of this methodology are discussed with respect to identifying successful learning activities in education. Such temporal predictions can be useful for practitioners in selecting learning activities that are likely to result in student success, which in turn can increase academic learning time. ALT, in turn, is associated with student learning achievement.

SUMMARY

The Dilemma: Qualitative vs. Quantitative Methodologies

Research methods in education used for much of the 20th century were largely quantitative methods. Experimental and quasi-experimental designs were commonplace (e.g., Campbell & Stanley, 1966), and analytical techniques included ANOVA, regression analysis and their extensions (i.e., discriminate, factor, canonical and path analysis). The basic problem is that this general linear models approach seldom yielded findings that could be directly linked to educational practice. Within-group and within-person variance was
often large, typically obfuscating differences between groups that could be attributed to so-called treatments, practices or programs (Medley, 1977; 1979). Cronbach & Snow (1977) further extended ANOVA to deal with aptitude-treatment interactions (ATI), with hopes of reducing the within-group variance. But this, too, was seldom successful in yielding significant results.

In the 1970s and 80s others explored alternative approaches that later became known as qualitative and case study methodology (cf. Guba & Lincoln, 1985; Stake, 1995; Yin, 2003). Qualitative methods have become widely used in educational research in the past two decades. One clear advantage of qualitative methods is that rich details of individual cases can give readers helpful insight into and understanding of the educational phenomena investigated. The unavoidable dilemma that often accompanies this approach is lack of justification for generalizability of findings. When samples are purposive and small, generalizability in the sense of making inferences from sample to population is seriously compromised. Indeed, respected qualitative methodologists avoid the term ‘generalizability’ and instead employ the notion of ‘transfer’ – i.e., results of what was found in this particular investigation may transfer to other similar situations the reader encounters (cf. Merriam, 1997).

To the great dismay of many researchers who use qualitative methods, federal agencies now are funding research with experimental designs and randomized trials that would be expected (on the surface at least) to create reliable and generalizable knowledge of education (cf. Shavelson & Town, 2002). There is no good reason to believe that history will not repeat itself here – the inescapable problem is within-group and within-person
variance. Human behavior is purposive and difficult to predict using deterministic methods such as the linear models approach.

A Middle Ground: Analysis of Patterns in Time (APT)

Frick (1990) proposed an analytic-measurement procedure called Analysis of Patterns in Time (APT). This is a paradigm shift in thinking for quantitative methodologists steeped in the linear models tradition and the measurement theory it depends on (cf. Kuhn, 1962). The fundamental difference is that the linear models approach relates independent measures through a mathematical function and treats deviation as error variance, but APT measures a relation directly by counting occurrences of when a temporal pattern is true or false in observational data. Linear models relate the measures; APT measures the relation.

For example, in an observational study of mildly handicapped children in elementary school, Frick (1990) found that, regardless of classroom context, when direct instruction was occurring (asking questions, providing explanations, demonstrating, giving feedback, prompting), these students were engaged on average about 97 percent of the time. In the absence of direct instruction, their engagement was about 57 percent. In other words, such students were 13 times more likely to be off-task during non-direct instruction. APT measures the temporal relation between direct instruction and student engagement.

This kind of APT finding is similar to epidemiological findings in medicine. For example, heavy cigarette smokers are 5-10 times more likely to have lung cancer later in their lives, and if they quit smoking the likelihood decreases (Kumar, et al., 2005). While causal conclusions cannot be made in the absence of controlled experiments, nonetheless one can make a practical decision to refrain from smoking and avoid smoky environments.
Patterns of Mode Errors in Human-Computer Interfaces

This recent empirical study illustrates the value of the merging of qualitative methods and quantification via APT. The present authors investigated conditions of mode errors when people use modern software.

Mode errors occur when the same user action results in more than one outcome, depending on the context. Mode errors can cause serious problems for software users, such as inadvertent destruction of important work, decreased productivity, and task incompletion.

Sixteen college students were each asked to perform eight computer tasks during usability tests of three modern direct-manipulation software interfaces. Stimulated recall interviews were conducted immediately afterwards as subjects watched themselves on videotape to clarify why they took certain actions during the tests. An observation system was devised for coding tapes of usability tests to record behavioral patterns of the participants.

Qualitative analysis of the results indicated three major types of mode errors: A) right action, wrong result; B) it isn’t there where I need it; and C) it isn’t there at all. A source of error analysis revealed that mode errors appear to result from eight categories of design incongruity: unaffordance, invisibility, misled expectation, unmet expectation, mismatched expectation, inconsistency, unmemorability, and over-automation. Consequences of mode errors included: can’t find hidden function, can’t find unavailable function, false success, stuck performance, inhibited performance, and inefficient performance.
Analysis of patterns in time (APT) was used as a quantitative method to determine the likelihoods of temporal patterns of types, sources and consequences of mode errors. Results of queries regarding temporal patterns were as follows:

<table>
<thead>
<tr>
<th>Query</th>
<th>Frequency</th>
<th>Likelihood (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>For type A ‘right action, wrong result’ mode errors (p = 34/52 = 0.65)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) If source is unaffordance, then consequence is can’t find hidden function or false success?</td>
<td>15 out of 34</td>
<td>0.44</td>
</tr>
<tr>
<td>then consequence is can’t find hidden function or false success?</td>
<td>10 out of 15</td>
<td>0.67</td>
</tr>
<tr>
<td>b) If source is invisibility, then consequence is stuck performance?</td>
<td>6 out of 34</td>
<td>0.18</td>
</tr>
<tr>
<td>then consequence is stuck performance?</td>
<td>5 out of 6</td>
<td>0.83</td>
</tr>
<tr>
<td>c) If source is misled expectation, then consequence is false success?</td>
<td>7 out of 34</td>
<td>0.21</td>
</tr>
<tr>
<td>then consequence is false success?</td>
<td>6 out of 7</td>
<td>0.86</td>
</tr>
<tr>
<td><strong>For type B ‘it isn’t there where I need it’ mode errors (p = 8/52 = 0.15)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d) If source is mismatched expectation, then consequence is can’t find hidden function?</td>
<td>8 out of 8</td>
<td>1.00</td>
</tr>
<tr>
<td>then consequence is can’t find hidden function?</td>
<td>8 out of 8</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>For type C ‘it isn’t there at all’ mode errors (p = 10/52 = 0.19)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e) If source is unmet expectation, then consequence is can’t find unavailable function?</td>
<td>10 out of 10</td>
<td>1.00</td>
</tr>
<tr>
<td>then consequence is can’t find unavailable function?</td>
<td>10 out of 10</td>
<td>1.00</td>
</tr>
<tr>
<td>f) If source is unaffordance, then source is unmet expectation, then consequence is can’t find unavailable function?</td>
<td>3 out of 10</td>
<td>0.30</td>
</tr>
<tr>
<td>then source is unmet expectation, then consequence is can’t find unavailable function?</td>
<td>3 out of 3</td>
<td>1.00</td>
</tr>
</tbody>
</table>

It can be seen above, for example, that when users did the right action but got the wrong result, the source of user error was often software interface elements that lacked affordance (functionality that is not obvious). This frequently resulted in users being unable to find a software function that was hidden from view, or they thought they did the task correctly only to find out later they had not (false success). When these conditions for
modes occurred, software users in this study were unsuccessful 67 percent of time in tasks they were trying to do. While a larger random sample of a broader range of users beyond college undergraduates would be needed to increase generalizability, these findings nonetheless illustrate measurement and analysis of sequential patterns of mode errors and their estimated likelihoods. These findings offer useful guidance to software designers who want to make their products easier to use by minimizing sequences of human errors such as those observed in this empirical study.

While this particular study does not examine patterns of teaching and learning that would be of more direct interest to educators, it does illustrate the value of a difference in approach to measurement. This study demonstrated the practical value of mixed methodologies where one first uses qualitative methods for gaining understanding of patterns and relationships, and then uses APT to code and quantify those temporal relationships in a manner that is useful to practitioners for making decisions based on APT predictions.

Further information on formal definitions of APT sequences and associated pattern counting rules can be found in Frick (1983), Chapters 2 and 5 (note that APT was originally called nonmetric temporal path analysis but the name was changed to prevent confusion with conventional path analysis; only the name has changed).

**Implications for Research in Education**

Now switch the frame of reference to learning in school. If we can determine that a particular activity in reading or mathematics is a good predictor of student success in that activity, then this is useful knowledge for practitioners. If these patterns are empirically verified through APT, practitioners are able to choose a series of curricular activities that
result in student success on a daily basis. Daily success promotes learning. Can we say whether any or all of these activities cause student learning success? No. But we could predict that successful student performance often follows in such activities.

APT does not verify cause-and-effect relationships, but instead verifies temporal relationships. APT can indicate the likelihood that $X$ precedes $Y$ precedes $Z$..., not whether $X$ causes $Y$ causes $Z$... For example, the event we call dawn is a good predictor of sunrise to follow. Dawn reliably precedes sunrise; it does not cause it. Nonetheless, knowledge of the predictable sequence can be useful for choosing a course of action.

The next implication is even more important. We do know from numerous studies of academic learning time (ALT), that the more often students experience success in the kinds of tasks that are measured on tests of learning achievement, the higher they are likely to score on such tests (e.g., Berliner, 1990; Carroll, 1963; Fisher, et al., 1978; 1980). Practitioners can select daily learning activities based on knowledge of this association: Choose daily learning activities that are compatible with curriculum goals (i.e., what we want students to learn) and which have the highest likelihood that students who engage in such activities will do them successfully.

Indeed, this is the approach that Maria Montessori took when developing her learning materials and activities (Montessori, 1965) – originally published in 1917, long before the ALT results were established. In developing learning tasks for young children to learn to read, write and do arithmetic, Montessori empirically determined which activities (called works) were most engaging by observing students who chose them from a wide range of works available in the classroom. When a student chose a work, Montessori assumed that it was meeting a student’s learning need at that time. She observed to what extent it held the
child’s attention, and whether the child performed the work successfully. Indeed, sometimes she tried to distract a student from a work by creating a competing activity nearby (e.g., students singing). If the student persevered in the task despite the distraction, this was a further indicator of its holding power. Montessori would also observe whether this pattern held for other students. Through this process of selecting works that promote engagement, the curriculum that Montessori developed consists of designed learning materials and activities that best held students’ attention, that they chose repeatedly, and that they performed successfully most often. These works are part of a Montessori classroom learning environment, where guidance of student learning occurs through student engagement with those works.

In summary, analysis of patterns in time can avoid the limitations of linear models and incorporate the insights of qualitative approaches. APT can help to identify conditions under which sequences of events are temporally predictable. Knowledge of such patterns can be useful for software developers and educational practitioners who want their products and learning tasks to result in successful performance day in and day out.
References


