Aptitude-Treatment Interaction as a Framework for Research on Individual Differences in Psychotherapy

Richard E. Snow
Stanford University

Aptitude-treatment interaction (ATI) methods are designed to take individual differences into account systematically in treatment evaluation. This article reviews the general concepts of aptitude and ATI and summarizes lessons learned in ATI research on educational treatments that should help ATI research on psychotherapeutic treatments. Recommendations for research design and data analysis address problems of aptitude distributions, multivariate aptitude complexes, detective work with scatterplots, disattenuation, treatment and therapist characteristics, therapist-client matching, ecological validity, outcome variables, statistical power, aggregation, and person independence. Example studies and hypotheses about the nature of ATI processes are also included.

In many fields of psychological, social, educational, and medical science, treatments are designed as interventions to achieve some individual or common good for human beings. When alternative treatments aimed at the same goal are available, the question is, Which treatment is best? Even when one treatment is adopted, a continuing question is, How can this treatment be made better? Because the persons treated usually are observed to differ in their response to treatment, and also to differ from one another in many other correlated ways, an important addition to these questions is, . . . best or better for whom, when, and why?

The aptitude–treatment interaction (or ATI) paradigm was invented to address these questions in consort. ATI methodology is designed to take individual differences among treated persons into account systematically in treatment evaluation—to assess the degree to which alternative treatments have different effects as a function of person characteristics and thus determine whether particular treatments can be chosen or adapted to fit particular persons optimally. Beyond methods for assessing interactions among person and situation variables, however, the approach offers a framework for new theories of aptitude interpreted as personal readiness to profit from particular treatment situations.

This article summarizes lessons learned in ATI research on educational treatments that should help advance ATI research on psychotherapeutic treatments. Dance and Neufeld (1988) have discussed the basic principles of ATI method with examples from psychotherapy research, so these need not be reiterated here. But the extensive experience with ATI in educational research suggests some clarifications and extensions of their discussion as well as some further recommendations.

I gratefully acknowledge the helpful comments of two anonymous reviewers on a draft of this article.

Correspondence concerning this article should be addressed to Richard E. Snow, School of Education, Stanford University, Stanford, California 94305.

Background

Terms and Concepts

Because there is misunderstanding abroad about ATI, one first needs to be clear on terms and concepts. The usage recommended here is as follows:

Aptitude should refer to any measurable person characteristic hypothesized to be propaedeutic to successful goal achievement in the treatment(s) studied; propaedeutic means needed as preparation for response to treatment. In other words, individuals differ in their readiness to profit from a particular treatment at a particular time; aptitude constructs are theoretical concepts fashioned to interpret these observed differences in person–situation interaction terms. An aptitude, then, is a complex of personal characteristics identified before and during treatment that accounts for a person's end state after a particular treatment.

It is important to note that this usage of the term aptitude differs from that common in English-language psychology through much of this century; that is, as here defined the domain of aptitude is not limited to intelligence or some fixed list of differential abilities but includes personality and motivational differences along with styles, attitudes, and beliefs as well. Also, no particular theory or measurement model for personality or ability is implied. According to Dance and Neufeld (1988), an ATI approach assumes a trait model in which aptitudes are presumed both stable and continuous. But there are no such requirements. It is happenstance that most aptitude constructs studied so far have been represented by tests or questionnaires built on classical psychometric models; although such models may use the term trait as shorthand and may impose continuous numerical scales, most apply without assuming fixed or continuous traits in any substantive sense. ATI research does seek complementary person and treatment characteristics that seem relatively stable, because characteristics that are easily changed pose no lasting problems. And treating aptitude variables as continua has certain statistical advantages.
But modifiability and continuity of aptitude differences represent questions for research, not assumptions; most ATI research has not yet examined aptitude change as a function of treatment, or the interrelation of traitlike and statelike properties of aptitude, or the multivariate distributional character of aptitude constructs. These seem to be especially important questions for further ATI research in psychotherapy.

In promoting such research, the ATI approach aims at new kinds of aptitude theories. The concept of aptitude recommended here returns to the original European meaning, emphasizing adaptation and the mutual fit (or misfit) of person and situation. An aptitude is thus a relational construct, interpreting the behavior of person-in-situation, and characteristics of the situation are as much a part of the definition of a particular aptitude construct as are characteristics of the person (see Snow, in press-a, in press-b, for discussion of the relational and situational character of this aptitude concept).

Aptitude constructs. These seem to be especially important characteristics of aptitude, or the multivariate distributional character of aptitude constructs. These seem to be especially important questions for further ATI research in psychotherapy.

A Simple Example

ATI is said to be present when for some group of persons an aptitude variable shows a different relation to an outcome variable in one treatment than it does in another. A typical study of instruction might contrast a high structure treatment (more didactic, direct, and teacher-centered) with a low structure treatment (more inductive, indirect, and student-centered). The outcome variable might be learning after some weeks in treatment as reflected by an achievement posttest. With a cognitive ability or achievement pretest as the aptitude variable, it is often found that the pretest–posttest relation is stronger in low structure than in high structure treatments. It appears that the completion and external control provided by high structure instruction helps initially less able students, reducing the difference between them and initially more able students, whereas low structure treatment accentuates this initial student difference. In the extreme, low structure is clearly the best treatment for high ability students, whereas high structure is clearly best for low ability students. A related pattern of results is often seen when these two instructional treatments are contrasted using test anxiety as aptitude. Highly anxious students do better with externally imposed structure; it presumably affords control of attention toward the task and away from self-doubts and worries. Low structure treatment is bad for anxious students, perhaps because it requires self-control of task attention; non-anxious students can provide this for themselves and so do well without high external structure.

There may be parallels connecting ability, anxiety, and treatment structure in psychotherapy as well. As argued above and through much of this article, however, these sorts of simple findings need deeper analysis and qualification in many ways before their use in practice. Often ATI effects signal the need for programmatic research on aptitude and treatment constructs and their improved assessment, not conclusions ready for application.

History and References

ATI hypotheses can be seen in ancient Chinese and Hebrew writings on education, in early Greek and Roman teachings, and in some European educational philosophies across the centuries (Snow, 1982). Binet invented the intelligence scale to serve interactionist purposes. But the modern definition of ATI for psychology stems from Cronbach (1957). Thereafter, a substantial ATI research program developed in instructional psychology (see Cronbach, 1967, 1975; Cronbach & Snow, 1977; Snow, 1977b, 1987, 1989a, 1989b; Snow & Lohman, 1984). More recently, the concept of aptitude as here defined, the methodology of aptitude research, and the requirements for building aptitude theories have begun to be studied in relation to the needs not only of educational but also of industrial and clinical psychology. To save space, these developments are not reviewed here, and background citations are given only for special references other than those listed above.

In passing, however, it is worth noting that some of Cronbach’s earliest ideas about ATI were derived in thinking about research on psychotherapy (see Edwards & Cronbach, 1952). Some related methodological suggestions (Cronbach, 1953, 1958; Cronbach & Gleser, 1953) also addressed the complexities of research on person variables in this setting. As current research addresses these issues anew, it needs to learn from the early work in psychotherapy as well as from experience in education.

Some General Admonitions

Three general warnings provide initiation for the specific recommendations offered later. These may guide thinking about problems in ATI research on psychotherapy that cannot be anticipated here.

The limits of simplification. For both research and exo-
tory purposes, all science must simplify. Especially in a new field, however, investigators must frequently step outside of their simple models and methods to identify complications and guard against distortions that might mislead or foreclose a promising line of inquiry. Much ATI research in education has certainly been limited by too simple a view of the phenomenon.

This exposition mainly considers simple experiments in which two or three treatments (\(T_x, T_y, \text{ and } T_z\)) are compared on a single outcome measure (\(O\)) and a primary aptitude variable (\(A_x\)). But secondary aptitudes (\(A_y\)) are brought in at times to show complications and to clarify lessons to be learned from them. Indeed, the recognition that aptitude variables, in particular, should not be considered only one at a time is an important first lesson. The world of person characteristics abounds in correlations, and it is unlikely that one aptitude effect is isolated from others. Every research design involves multiple aptitudes and higher order interactions whether it includes them formally or not.

Similarly, treatment and outcome variables have hidden correlates that can moderate results across contexts and thus across studies. In research in education and also in psychotherapy, one must take a more complex view of replication (and of meta-analysis) than that typically taken in laboratory psychology (where troublesome interactions also occur across contexts). Some complications involving multiple treatment and outcome variables, and related difficulties, are addressed in later sections. The overarching point is that each experiment is a case study, limited to its time, place, and human constituents. The best way to understand its results and to link them to the results of other studies is to obtain the richest possible description of its context and process and to limit simple generalizations therefrom (see Cronbach, 1982a, 1982b, 1986).

The limits of statistical significance testing. A different sort of simplification limits and distorts progress by automatically removing from further consideration ATI effects that fail to reach some conventional level of statistical significance. Certainly investigators need a decision rule to focus attention. But \(p = .05\) does not divide ATI effects into two classes: reportable and zero (i.e., unreportable). And a table of \(p\) values or \(F\) ratios or asterisks is not a report of results. Every ATI report should provide descriptive statistics within treatments both for results judged significant and for those judged nonsignificant. Consistent nonsignificant trends are at least as valuable for the purposes of further research as are incoherent significant results. And estimates of relationships including confidence intervals around them are more valuable than null hypothesis tests for these purposes.

A huge amount of ATI research in education has been wasted through failure to recognize the primacy of statistical power considerations in significance testing. In other words, Type I errors have been avoided at the expense of what may be a vast number of Type II errors. Investigators need to decide what size effect they want to be able to detect and what power level is needed to do so. Inasmuch as most ATI studies cannot practically obtain the sample sizes needed to achieve adequately powerful tests, considering and reporting the actual power of a study to detect ATI (alongside the results obtained) helps to avoid simple null conclusions and to improve research design for future tests. A number of lessons for research design flow from the power issue.

The limits of existing methods. There is also a natural tendency to simplify by allowing existing methods, rather than a conceptualization of the substantive phenomenon, to dictate the course of research. Ready-made aptitude measures and comparison treatments can be gotten off the shelf for trial. The data can be fitted into canned statistical analyses. But the kinds of aptitude and treatment contrasts that derive from earlier theoretical traditions may not be the best candidates for ATI research in psychotherapy. A huge amount of research effort in education has also been wasted on relatively blind searches through the catalogue of existing variables.

There may be existing aptitude and treatment constructs in psychotherapy that are ripe for ATI research. But there are also likely to be new conceptions of aptitude and treatment derivable from thinking about persons-in-situations and analyzing cases in that light, without reliance on existing nomenclature. In either case, there is no substitute for careful, deeply substantive consideration before aptitude measures are adopted and treatment contrasts are designed.

Recommendations on Basic Designs and Analysis

This section offers more specific recommendations regarding basic designs and data analysis for ATI research. A following section takes up some advanced problems. Because space is limited, investigators should consult the book-length presentation (Cronbach & Snow, 1977) for full explication on most points.

Research Designs

Four basic ATI designs may serve useful purposes in psychotherapy research. There are also various complex designs that build on these but are not taken up here.

Standard design. A simple randomized between-persons design and hypothetical result akin to that discussed by Dance and Neufeld (1988) is shown in Figure 1a. The regressions of \(O\) on \(A_x\) differ for three treatments (\(T_x, T_y, \text{ and } T_z\)). To optimize outcome in practice, one would apply \(T_x\) to persons above Point \(x\), and \(T_y\) to persons below Point \(x\), on \(A_x\). If the treatments differ in cost, decisions may change. For example, if \(T_y\) is less expensive than \(T_x\) and not much different in effect for persons above Point \(x\), then it might be used instead of \(T_x\): this changes the treatment choice point from Point \(x\) to Point \(y\) on the aptitude continuum. There are formal methods for converting \(O\) variables to a utility scale, thus transforming the picture (see Cronbach & Gleser, 1965). But note that this conversion can either make or break practical ATI considerations. For example, if \(T_x\) is least expensive, the conversion shifts the \(T_x\) regression line upward in Figure 1a, perhaps making it the treatment of choice for all persons in the sample.

Cost analysis of particular existing therapies is an important but secondary issue; the first question for research is why particular ATIs occur as they do. Existing evidence from individual studies needs to be pursued more deeply, through case-by-case analysis. If theoretical interest attaches mainly to a particular treatment contrast, then attention should focus on clues in the
case analysis that suggest what other aptitudes might be brought into subsequent studies. By selecting aptitudes that then exhibit ATI, one demonstrates that the treatment contrast is to that degree understood. If the theoretical focus is on the meaning of a particular aptitude construct, then case analysis may suggest what treatment manipulations in further studies might clarify this meaning further. Experimental manipulation of ATI shows that the aptitude construct is to that degree understood. One can start from either focus or both jointly. And the analysis and classification of accumulated earlier cases in some research program, practice, or clinic could use ATI thinking to generate important research hypotheses even where aptitude and treatment variables have not been systematically studied in interaction (see Kanfer & Phillips, 1970, on this recommendation).

Treatment revision design. One variant examines a single treatment and attempts to improve it with respect to some aptitude, as pictured in Figure 1b. The between-persons version of this design was first proposed by Bunderson (1969) and used in formative evaluations of computerized instructional programs. The regression of O on A for T is a starting point, obtained perhaps in an initial correlational exploration of T effects. The investigator then uses clues from cases low in the O-on-A distribution to guide revisions, thus creating T. Suppose a second study with new persons then produces the T regression pictured. Another round of analysis and revision and a third study might produce the T regression pictured. The aim is to reach optimum outcome by adapting treatment to remove the negative effects of low A over a series of correlational studies (that add up to an experiment).

A threat to this result observed in educational studies is suggested by the dashed regression slope. Treatment revisions that improve outcome for low A persons may reduce its effectiveness for high A persons, who were initially well treated. This may suggest that something important for high A persons has been altered or removed in the treatment revision process and may also hint that changes have brought some A into relevance.

A within-person design may also be considered. Psychotherapy might conceivably be designed in multiple stages, so long as interim O measures exist. Each stage might then be a further adaptation of treatment with respect to A,. Mixed designs are also possible, where half-samples are randomly assigned to a new treatment stage or retained in the initial stage.

Aptitude growth design. In another variant, A is itself the target of treatment, so O is equivalent to A', as in Figure 1c. T appears best for persons initially low in A.. After a period of growth in this treatment, however, other treatments may become optimal for continued progress; a shift at Point x to a transitional T and another shift at Point y to the finishing T can be imagined.

The evaluation of such schemes may require multistage experiments as in treatment revision designs. Also, when many A or O assessments are available in a time series, other forms of analysis may be clearer and more powerful than the standard regression analysis. Rogosa (in press) has demonstrated a growth curve approach to the study of ATI in such situations.

Regression discontinuity design. A variant that may be useful in evaluating ATI across certain other patient situations uses
a regression discontinuity approach. If persons are already assigned to different treatments on the basis of $A_1$ (or some correlate), then evidence for ATI comes from the degree to which a common regression line does not fit the full data set across Point x. Two kinds of discontinuity are suggested in Figure 1d. The contrast of $T_A$ and $T_B$ suggests a main effect; the two regression lines are parallel, with that for $T_A$ shifted upward or that for $T_B$ shifted downward for all persons so treated. The contrast of $T_A$ and $T_C$ suggests ATI; if extrapolated, the differing slopes would intersect. Although extrapolations are tenuous, such results at least suggest hypotheses worth further examination.

Suppose low standing on $A_1$ (below Point x) is used in deciding to institutionalize persons for treatment, but the same therapy is applied to both inpatients and outpatients. The contrast between the $T_A$ and $T_C$ slopes in Figure 1d might be taken to indicate the depressing effect of institutionalization on therapeutic progress. On the other hand, the contrast between the $T_A$ and $T_B$ slopes might suggest a beneficial effect of the institutional context coupled with therapy. Obviously, regression discontinuity can also be used to compare different therapies across institutional or other context lines.

Despite the amount of prior research on ATI in education, almost all examples use one or another form of standard design. One might expect that instructional development work would make use of the treatment revision design, but there are few examples actually carried through the series. Despite much expenditure on treatments that work directly to develop intellectual skills for learning, evaluation studies have generally not used aptitude growth designs. Regression discontinuity is an obvious way to evaluate placements into special education programs for retarded, learning disabled, and gifted students, but again there are no good examples. Research in education has really not yet gone beyond the question of whether understandable and useful ATI can be demonstrated. The shortcomings in much of this work lead to the suggestions for research on psychotherapy laid out below.

**Preservation of Meaning in Aptitude Variables**

A first goal is to preserve the substantive meaning of the aptitude construct by choosing appropriate models of measurement and statistical analysis. This end is served by several recommendations.

**Examine aptitude distributions.** Assumptions about continuous, normal, linear distributions should not be adopted blindly. Examination of obtained score distributions may suggest (and has at times suggested) discontinuities that may have substantive meaning to be captured by forming discrete groups. Bimodal distributions obviously suggest this. Also, distributions that are continuous and linear in one population or region thereof may show anomalies in other populations or regions. On some continua, extreme scores may identify qualitatively different psychological groups. And some aptitudes studied are at base discrete (e.g., certain style constructs; sex as a proxy for psychological differences). Dance and Neufeld (1988) have rightly noted that aptitude variables that readily distribute into discrete classes should be analyzed and interpreted as such (see Gangestad & Snyder, 1985). Several discrete class measurement models are available for use in ATI work (see, e.g., Haertel, 1984; Rindskopf, 1987).

**Avoid forced categorization.** But continuous measures should be analyzed as such. Regression methods provide estimates of strength and form of relationships in the most powerful way. Splitting a distribution (e.g., at the median) inflates error variance by failing to use the discriminating power of the continuum. Trichotomies or other forced levelings are worse because they also ignore (at least in conventional analysis of variance [ANOVA]) the ordinality of the continuum. Furthermore, forced splits are arbitrary and sample-bound, so they confuse attempts at replication (e.g., above-median anxiety in one sample may be below-median anxiety in another).

**Use norms where possible.** If population norms exist for an aptitude measure, interpretation in a sample should be referenced to them. This adds meaning to any one study and locates each study in its region of the population so that replications can be meaningfully compared.

**Avoid standardization of scores in a sample.** Standardizing scores has advantages but is hazardous in research on a sample not referenced to norms. Particularly to be avoided is standardization within treatments. This common practice erases from view variance differences in outcome resulting from treatment effects. In other words, regression coefficients, not correlations, should be used to compare treatments. Raw regressions are usually best. For some purposes, measurement scales for either aptitudes or outcomes can be transformed to add meaning, but never separately within treatments.

**Beware confounded extreme groups.** Extreme groups design fits ANOVA and provides a powerful statistical test, but choosing such groups wisely is difficult. Especially in a new research program, one rarely has the knowledge needed to do this effectively. One must assume linearity and an isolated aptitude. When a second aptitude is moderately correlated with the first, confounding occurs in proportion to the degree of correlation. As shown in Figure 2a, the shaded extreme groups on $A_1$ differ also on $A_2$, thus clouding interpretation. $A_2$ need not be measured to cause this confounding; it exists as variance in the sample. If it is measured, then extreme groups might be defined to include it, but this involves sampling persons in more complex ways, using the bivariate contours. Extreme groups design also seems to fit psychotherapy research poorly. Persons can be assigned randomly to treatment, but then all comers are usually treated. Throwing away the midrange data serves no good end.

**Include a second aptitude explicitly.** Theoretical interest often attaches to one aptitude construct in relation to one treatment contrast, but including a second (or third) aptitude has advantages and is usually inexpensive. $A_2$ might be chosen for its close relation to primary $A$ and $T$ variables in a conceptual network or for its importance as a moderator in psychology generally. Figure 2b shows the effects of higher order ATI when both ability ($A_1$) and test anxiety ($A_2$) are used as aptitudes to contrast highly structured teaching ($T_A$) with unstructured student-centered teaching ($T_B$) as treatments. To avoid a three-dimensional picture, Figure 2b identifies regions of the bivariate aptitude space labeled to show which treatment was best in each. It shows that structured treatment was best for students high in both ability and anxiety or low in both; unstructured treatment was best for students showing high-low, low-high,
and middle level profiles. The results replicated. However, the aptitude-outcome relations may have been curvilinear in these studies; sample size was insufficient to test this. In any event, these results are seen to complicate considerably the simple example of ATI given early in this article.

**Inspect relationships in scatterplots.** An important rule is not to allow packaged computer analyses to dictate the results. Scatterplots should be carefully inspected. Figure 2c shows some reasons why. When one attends only to regression slopes, a striking ATI appears. However, the white dot group may have one and possibly two outliers; its regression slope would be much less steep without these points. One does not drop outliers purely on statistical grounds; each is a real person. Rather, one studies the cases to find plausible reasons why they are where they are in the scatterplot. But recognizing the influence one or two points may exert in a small-sample study and checking out that influence will temper otherwise enthusiastic conclusions.

Two other notable features of Figure 2c are the restricted range on O exhibited by the black dot group and the apparent bimodality of both groups on A1. Treatment can make outcome variance larger or smaller, so this further emphasizes that correlations within treatments are not the statistics to look at. The bimodality is a question for further inquiry. If A1 really is best treated as discrete, other information from case analysis should justify this, and further research can be planned accordingly.

The data of Figure 2c actually come from a memory experiment using college students in which scholastic ability scores served as A1. Hence, bimodality is probably peculiar to this sample. Also, there is only one treatment here; the black dots are men and the white dots are women. This further underscores the need to include secondary A variables; with ability as A1, sex might be chosen as A2 to index whatever other psychological differences attach to gender. No matter what else one concludes here, results for men and women differ enough to justify further inquiry.

A second reason to plot data person by person and to consider what else is known about each. These data come from a study of students working in an innovative treatment based on small group problem solving; A1 is verbal ability and O is knowledge acquisition after some weeks. The solid regression line shows the overall relation of O to A1.

The scatterplot displays two features that deserve further detective work: There is a suggestion of curvilinearity, unusual for an ability–learning relation, and the points are densest in the middle to upper region of A1. The plot looks as if one bivariate distribution is superimposed on another with AO relations running in contrary directions in each to reduce the overall linear regression. One could fit a single curvilinear function to these data, but an adaptation of the off-quadrant analysis proposed by Marks (1964) was considered simpler and more useful in this instance. Other aptitude dimensions, available from a personality inventory administered at pretest, were used to search for some A2 that might yield the dashed regression line in Figure 2d (i.e., that would distinguish cases above the solid regression line from those below it in the middle to upper region of A1). It turned out that students in this region below the solid regression line were on average more independent, achievement-motivated, and task-oriented, and less altruistic and interperson-
ally oriented, whereas students in this region above the solid regression line showed the opposite pattern. A corresponding to the dashed line might thus be an aptitude complex roughly described as independent task motivation versus interpersonal orientation and mildly correlated with ability in this middle to high range. The resulting hypothesis is that able students who are also highly motivated and oriented toward independent work do poorly in this treatment because it demands cooperative interpersonal activity in small groups, but able students who value interpersonal activity more than independent task activity do well in this treatment. A treatment that reduced or eliminated group work in favor of individual learning would be hypothesized to reverse this relationship.

The point is that scatterplot and case analysis within treatments can yield important advances in ATI research on particular treatments. Although one aptitude may be of primary interest, carrying collateral aptitude information along in a study to facilitate this sort of detective work can yield valuable payoffs.

**Disattenuate relationships.** Observed regression slopes that are marginally different can become markedly different when corrected for measurement error in A or in both A and O. Research is properly interested in theoretical relations, so correcting both A and O for error is appropriate. Personality measures are often not highly reliable, so the dictum to disattenuate O-A relationships should carry special force for psychotherapy research. Reasonable estimates of reliability for each measure are of course required.

**Preservation of Meaning in Treatment and Outcome Variables**

Relative to A variables, psychotherapy research has accumulated much more experience in considering T and O variables, so less needs to be said about them here. With ATI the issue, however, some lessons from educational research deserve note.

Describe treatments as fully as possible. Labels do not a treatment make. When educational treatments that appear superficially similar have produced unexpected results, a failure of replication has often been announced. But a closer look at details sometimes yields explanation. Also, the treatment intended may differ from the one implemented. Thick description of the actual treatment is needed, and it should include characteristics of the situational contexts in which treatments are embedded. One needs to be able to trace varying results to other interactions that may be subtly varying across contexts. In psychotherapy research, the production and use of detailed treatment manuals should be standard operating procedure; it is a major step in this connection (see Smith & Sechrest, 1991).

Describe therapist characteristics and styles. Teachers may or may not follow treatment prescriptions exactly, and their varying natural styles may fit one treatment better than another. In one study of 39 classrooms, for example, within-class correlations between measures of student test anxiety and mathematics achievement ranged from +.36 to −.81 (Helmke, 1988). The immense range suggests that teacher style was an important moderator; in effect, the teachers were administering different treatments. Therapists clearly also vary in this way (see Beutler, 1991). Thus treatment description needs to include therapist characteristics and styles.

**Strive for ecological validity in experiments.** Considerations of Treatment × Therapist interaction suggest multifaceted experiments in which therapists are scripted to impose different manualized treatments for randomly different clients. Perhaps therapists can also be chosen to represent different levels of some characteristic. Obviously other treatment variables can also be crossed with these. But the ecological validity of the resulting mixtures of conditions is an important issue.

In some educational studies, teachers have been scripted to vary treatments systematically across multiple classes. Tutors have been trained to do this for different individual students. But careful analysis of transcripts sometimes shows that all teachers do not work equally well with all treatments. Crossing teachers and treatments in a multifaceted design can also produce awkward unnatural conditions for some teachers and consequent artificial effects in the results. After all, neither teachers nor therapists are professional actors, and even actors do not fit all available scripts. Thus, the fit of therapist and treatment needs to be considered; degrees and kinds of misfit can be subtle. For this reason also, the use of pseudotherapists (also pseudoteachers) is not recommended.

The methodological advice here is to sacrifice crossed design in favor of nested design wherever available therapists seem to fit some treatments naturally, and not others. Ecological validity and the trade-offs between external and internal validity of experiments cannot be addressed here (see Cook & Campbell, 1979; Cronbach, 1982; Snow, 1974). But it is clear that experimental contrasts need to be made as representative of real practice as possible if results of psychotherapy research are actually to be applied. It also seems clear that further research needs to focus on the problem of training therapists (as well as teachers) to be competent in choosing and using the range of treatments that different clients match.

Analyze therapist–client match. If therapists can be fitted to treatments, why not also to clients? Therapist–client matching in psychotherapy has been studied for some years (see Beutler, 1981; Chartier, 1971; Kelly, 1990; Razin, 1971). But this line of research may not yet have fully recognized the complexities involved.

In education there is evidence that some teacher–student matches work well and others poorly (see, e.g., Brophy & Evertson, 1981). However, much too simple a view has been taken of what constitutes a beneficial match and how this can be detected. It may be found in interpersonal similarities in some instances and complementaries in others. Perhaps it must be traced back from outcomes through transactional processes during treatment and only then to therapist and client characteristics. The relevant characteristics may be both multivariate and different in therapist and client. Thorny methodological problems attend these questions that have yet to be adequately studied. At least one recommendation is to avoid dyadic indices in interpersonal matching: these often lead to artifacts and unnecessarily complex interpretations, which are revealed when analysis returns to the separate person dimensions (Cronbach, 1958).

Include perceptions of treatment and therapist. The studies of teacher style and student anxiety noted above replicate a frequent finding. Teachers who impose highly structured conditions reduce the negative effects of anxiety on learning; the
negative relations are stronger where teacher structure is not imposed. But some research that demonstrated this ATI experimentally also later showed the same result with a constant treatment that blended the two teaching styles; students were merely classified on their perceptions of treatment as high versus low structure (Dowaliby & Schumer, 1973). In other words, perception of teacher style alone turned the anxiety–learning relation one way or the other. Thus, client perception of treatment, as well as its actual characteristics, may often be an important interacting influence; there is some evidence from psychotherapy research that this is so (Garfield, 1986; Gurman, 1977a, 1977b). This possibility also complicates research on therapist–client matching.

Include all relevant outcome variables. To evaluate treatments, one obviously needs to include all important outcome variables. Educational studies have sometimes found different ATI effects for different O variables, thus providing insight into subtle treatment processes. Unfortunately, each result has usually been interpreted in isolation. In psychotherapy research, there are also examples of differing results for different outcome variables (see, e.g., Strupp & Hadley, 1977, 1979); here too, multiple kinds of side effects and recidivism are always special concerns.

Separate ATI findings should not be interpreted independently when O variables are correlated. Outcomes can sometimes be orthogonalized to clarify interpretation. They can also be examined jointly if converted to utilities or some other common scale, so that ATI effects on the difference between outcomes can be assessed. Judging from experience in educational research, multiple outcome analysis can enrich the understanding of ATI substantially if it is designed to preserve the meaning of each measure.

Avoid expressing outcomes as difference scores. The meaning of O variables is also preserved by avoiding arbitrary indices of “difference” or “change.” Although simple change measures are intuitively appealing, they are treacherous. Difference score metrics involve assumptions that are usually not met. Their use often befuddles rather than clarifies the analysis. Besides, if therapeutic process and outcome are multivariate, no one O variable should be singled out as special because it seems superficially similar to one A variable. Unless multiple time series measures are available, the recommendations of Cronbach and Furby (1970) should be observed (see also Beutler & Hamblin, 1986). A similar argument, along with the consideration of power, dismisses covariance analysis in most instances.

Some Advanced Topics

This section touches only a few of the advanced issues that may come up as ATI research in psychotherapy progresses. The aim is to anticipate likely complexities. Those noted are mainly extensions of problems already discussed.

Power, Aggregation, and Person Interdependence

Power versus detective work. Because statistical power is vital to detecting ATI, many of the above recommendations aim at enhancing it. But some—most notably the advice to entertain multivariate or curvilinear hypotheses—would seem to run counter to this emphasis, especially given the limits on sample size typical of psychotherapy research (and of much instructional research). There is usually a trade-off between improving power and exploring complexity. For a given sample size, simpler hypotheses are more powerfully tested than are complex ones. Design and analysis options that offer increased statistical power usually do so by limiting the kinds of ATI effects the experiment can detect. For example, choosing extreme groups or stratifying or matching on aptitude before assignment to treatment can increase the power of the analysis, thus reducing the risk of missing a simple effect. But this increases the risk of missing meaningful but more complex ATI. To see this, superimpose Figure 2a onto Figure 2b; an extreme groups design using only A1 would completely miss an important bivariate effect. Sometimes statistical as well as experimental controls can be used to reduce error terms and thus enhance the sensitivity of the analysis to one effect, but this also can lead to oversimplified, misleading conclusions. The alternative is to accept lower statistical power as the price of preserving or increasing the value of the data for exploratory detective work. What strategy is best depends on a host of considerations, including the stage of the research program and the strength of previous theory development, as well as the risks an investigator is willing to take in a particular situation. At the least, however, an investigator should always estimate the power needed to test each hypothesis before the study and report it alongside results.

Sometimes one can increase power with respect to a main hypothesis while still exploring many secondary issues. And at base, of course, one can increase both the power of an experiment and its value for exploration by increasing its sample size. Although an investigator must conduct a needed study with whatever sample can be practically obtained, programmatic research should try to build up large samples for each primary hypothesis. In education this is done by including multiple classes (and thus teachers) in each treatment, inasmuch as treatments cannot usually be experimentally manipulated within classrooms. In psychotherapy, it may also be possible to aggregate into a single analysis multiple sites (e.g., multiple therapists, clinics, or other health care institutions). But psychotherapy research has two options in this process. It can assign sites to treatments as in educational research, or it can conduct the same small experiment at each site, because individual clients or small groups can be assigned randomly to one or another treatment as they come. Consider the first strategy, then the second.

Between-groups versus within-group regressions. In the Helmke (1988) example noted earlier, 39 classrooms displayed a range of correlations between student anxiety and learning outcome, and this was associated with variation in teacher styles. Each class thus has a within-group regression of O on A (i.e., Helmke's within-class correlations need to be transformed to regression slopes using the respective standard deviations). But each class also contributes to a between-groups regression that relates class average O to class average A across the sample of classrooms. This between-groups regression can be very different from the individual class regressions, or their pooled regression, and both can differ from the regression that would be obtained for the sample of individual students with class mem-
bership ignored. In other words, ATI can occur at two levels. If in the Helmeke study classes differed in average anxiety, for example, the between-groups regression might be negative for low structure teaching styles, but within-class regressions might vary from positive to zero as class average anxiety increased. High structure classrooms might show the same progression of within-class relations as class average anxiety increased but a positive between-groups regression over the continuum of class averages. Furthermore, in correlational studies of this sort, teaching style differences might create an ATI pattern across the within-classroom regressions while at the same time being a result of class average differences in anxiety.

When sites are assigned to psychotherapeutic treatments, the lesson is that between-site and between-sites regressions have to be disentangled and compared within and between treatments. To predict each person's response to treatment, one must look at each person's A level in relation to the average A for the group or site. Testing overall treatment regressions for persons without regard to site confounds between-groups and within-group sources of ATI. Unfortunately, unless there are many sites and many persons within sites, both between-groups and within-group regression tests are weak.

The second strategy of conducting miniature experiments in many sites adds up to a more powerful regression comparison and is preferred wherever it is possible. But between-sites within-treatment regressions still must be studied: because persons are not assigned randomly to sites, the sites may differ in average A. One uses a pooled within-site regression to do this.

Group therapy. The above problems are exacerbated when persons receive treatment in groups within site. Because of social perception, comparison, and role-taking processes operating in groups, members of such groups are not independent. In educational research on small groups, it has been shown that effects can depend on the person's A, the mix of A in the group, and the role the person adopts in group interaction (Webb, 1982, 1983). Note also the effect of imposed group process on motivational style distinction between achievement via independent and interest and bringing in others as they are expected to moderate primary ATI or to be secondarily important in their own right, in some particular situation. If the list becomes long, one can sometimes reduce to composite A variables. In instructional research with cognitive abilities, for example, a common strategy extracts a general ability factor and then a differential factor to contrast relative strength in verbal versus spatial abilities; a battery of cognitive tests is thus reduced to two orthogonal A composites. With personality variables, however, there are often no strong theoretical or empirical bases for constructing composites. Also in most psychotherapy research as well as much educational research, sample sizes are too small to justify statistical definition of composites.

As multiple A variables and higher order ATI have been studied, the notion of aptitude complexes has emerged from the recognition that different aptitude combinations sometimes interact with the same treatment contrasts. Figure 2b gave an example of an Ability X Anxiety interaction with teacher structuring. The same studies and other research suggest that a motivational style distinction between achievement via independence and achievement via conformance interacts with the same treatment contrast. Thus, when one is using conventional ATI methods (even holding T and O to single variables), a complicated analysis and a large sample are needed to estimate a regression equation involving four A variables and their interactions. As more variables are added, the conventional analysis quickly becomes weak, impracticable, and probably uninterpretable.

But persons are whole beings, not simply lists of variables; in the above example, each is a case point in a four-dimensional A space. Might such a space be partitioned into regions wherein common effects of T on O can be expected? Such regions might define new complex aptitude constructs relevant to particular treatment situations. Figure 2b may outline a first crude stab at such constructs.

Using one prominent theory of achievement motivation, for example, both ability and anxiety can be interpreted as influencing arousal in achievement situations (as well as having particular cognitive effects). Perhaps high ability, high anxiety persons and low ability, low anxiety persons can be described as overaroused and underaroused, respectively, when experiencing low structure treatment. By not imposing external regulation on behavior, such treatments demand internal regulation, which neither type of person can provide. High structure treat-
ment imposes external regulation, which has the effect of moving both kinds of persons toward optimal conditions for learning by reducing or increasing arousal, respectively, relative to the low structure treatment that already provides optimal arousal conditions for intermediate ability-anxiety persons. Adding an index to contrast relative need for achievement through independent activity versus need for achievement via conformance to external structure moderates the regional boundaries of Figure 2b. The T_A region expands for persons needing independent activity and contracts for persons needing conformance to structure; perhaps some curvature of boundaries is also introduced, inasmuch as achievement via independence has been shown to correlate with ability. Choosing satisfactory construct interpretations for these regions would require much more case analysis, but note that this sketch is based on empirical examples. At least this meager first step has reduced four A variables to two and brought a T variable into the definition of aptitude in the Aptitude × Treatment regions that result.

To reach complex aptitude constructs, some investigators have applied forms of profile analysis to classify or match persons on aptitude profile similarity (or complementarity). This possibility deserves attention, but it brings in other problems. First, response to treatment ought to be part of the profile, as suggested above. Second, there are a number of different measures of similarity, each with somewhat different meanings and complications; as noted in a previous section, dyadic measures are particularly complicated and should be avoided. Finally, profile measures step even further away from the original, more directly interpreted A variables than do the regional partitionings exemplified above, especially when many A variables are combined to reduce profile complexity; there are attendant problems in estimating profile reliability (see Gleser & Crespo da Silva, in press).

Various typological constructs are also possible. The Type A-B distinction among psychotherapists (Chartier, 1971; Razin, 1971) is an old example. In some more recent research, there has been a return to typological descriptions to boil down multivariate complexes. The Type A personality for heart attack risk is a well-known example. The Type T, or sensation-seeking personality, is a more recent addition (Farley, 1989). Again, in any such construct, response to treatment(s) needs to be explicitly included if it is to guide research in psychotherapy.

New measurement models. There is a burgeoning new frontier in instructional psychology aimed at elaborating substantially the older, narrower conceptions of aptitude, learning—development, and achievement and devising new forms of assessment for the new constructs (see e.g., Snow, 1989c). Some work focuses on improving the diagnosis of learning progress (see, e.g., Frederiksen, Glaser, Lesgold, & Shafto, 1990). Some research addresses measurement methods that combine cognitive and motivational constructs (see, e.g., Kanfer, Ackerman, & Cudek, 1989). Some seeks to adapt or replace the assumptions of traditional psychometric models (see, e.g., Frederiksen, Mislevy, & Bejar, in press). Much of this work may be suggestive for improving both aptitude and progress assessment in psychotherapy. Some of these approaches connect particularly with the problem of assessing aptitude complexes. But none are yet sufficiently evaluated in educational research, much less in psychotherapy.

Some Formative Hypotheses

Methodological lessons are more likely to transfer from education to psychotherapy research than are particular substantive hypotheses, so method has been the main thrust here. But three general substantive themes can be seen in the ATI evidence to date that deserve brief note in conclusion; each offers a metalevel conceptualization that may help guide more specific hypotheses about ATI in psychotherapy.

Capitalization and Compensation Processes

Much evidence suggests that individual differences come into play as aptitudes as a function of situational demands and affordances. Therefore, these demand—affordance conditions can be manipulated in treatment design to capitalize on strengths in some aptitudes while at the same time compensating for weaknesses in others. A treatment that demands, or affords the opportunity to exercise, a particular personal capability or tendency can be made to use this characteristic as an asset in compensating—that is, in circumventing or remediating—some other personal characteristic. To return again to the previous example, perhaps teacher-imposed structure capitalizes on one kind of tendency—toward achieving via conformance to external structure, in order to compensate for another—instability to control the negative arousal effects of overly high or low anxiety. Removing teacher structure capitalizes on an opposing tendency—toward achieving via independent, personal control of ability and arousal.

Different therapeutic treatments also differ in both the degree and kinds of demand—affordance structures they impose. Analysis of treatment contrasts in these terms may help identify the aptitudes each calls into play and thus the kinds of capitalization—compensation processes that might be observed. A simple example is the form of systematic desensitization that requires clients to generate and control vivid visual imagery; it capitalizes on an ability that not all clients possess. A contrasting treatment will need to compensate for this inability, perhaps by using an opposing strength in verbal— analytic ability. A more complex example comes from Shoham-Salomons (1991); see also Shoham-Salomons, Avner, & Neeman, 1989) results with paradoxical interventions, which suggest how one treatment might sometimes work on both sides of an aptitude difference to reflect different capitalization—compensation processes. Persons who are reactance prone defy therapist directives to engage deliberately in the symptomatic behavior; they mobilize their resistance and thus become able to control the symptomatic behavior directly. Persons who are not reactance prone appear better off initially in self-control therapy; with paradoxical treatment, however, they follow the therapist directives and thereby gain increased self-efficacy with respect to controlling the symptoms eventually. Apparently then, paradoxical interventioncapitalizes on high reactance to reduce symptoms immediately; it also circumvents low reactance to build an alternative aptitude for reducing symptoms in the long run.
Zones of Tolerable Problematicity

Each treatment's demand-affordance structure needs to be understood, of course, in relation to the aptitude patterns of the persons being treated, and these require description in terms relative to populations, not just to samples at hand. In this connection, an emergent notion in ATI research posits that for each person-treatment combination there is a threshold of demand near which optimal learning progress can occur. Treatment demand that is far below this threshold elicits algorithmic unmotivated response and thus little progress. Treatment demand too far above threshold is overwhelming and aversive, eliciting maladaptive response, either algorithmic and rigid or helpless and random, and also little progress. In a zone of tolerable problematicity around this threshold, however, treatment demand engages heuristic psychological processes that enable the assembly and control of adaptive response and thus progress toward the goal.

Although this concept emerges from educational research on cognitive abilities in relation to learning and problem solving (Elshout, 1987), it has been used to interpret ATI results for achievement motivation and anxiety, such as those used as examples here (see Snow, 1989a). It suggests that an important future research goal should be to identify and understand these thresholds and zones for each type of person-treatment combination of interest. It also suggests that some important kinds of ATI may be curvilinear, especially when personality variables serve as aptitudes. The implications for psychotherapy research are the same. For each type of person and treatment situation there is likely to be a threshold or zone within which optimal effect is achieved and outside of which it is not. Research needs to focus on understanding this phenomenon.

Prototypical Person-Treatment Combinations

Discussion of aptitude complexes, demand-affordance structures in treatments, and key thresholds and zones in the person-treatment interface implies finally a different approach to conceptualizing ATI in psychotherapy that seems akin to Mischel's (1984) brand of person-situation interactionism. Not all person variables or treatment variables are relevant to all persons. Rather there may be bundles of person and treatment features that reflect prototypes of successful person-treatment outcomes. These prototypes may display local consistencies although they may not apply across all contexts. Conventional ATI research methods may be used to help identify aspects of these prototypes or bundles. So may case-by-case analysis and accumulation in one site. But the resulting ATI constructs will be local theories, rich descriptions useful for understanding and using ATI in a particular time and place, with a particular range of persons (Cronbach, 1975, 1982b; Snow, 1977a). As persons or situations depart from the prototype, this local theory no longer applies, although some other local theory might. The result is a loose confederation of miniature prescriptions that may even be in some ways inconsistent with one another; it is not a unified grand design for psychotherapy. But such local theories may prove useful whether or not general theories ever prove possible in the ATI field.

References


Farley, F. (1989, August). *The Type T personality goes to school*. Invited
address at the 97th Annual Convention of the American Psychological Association, New Orleans.


