Empirically Based Myths: Astrology, Biorhythms, and ATIs*

Ronald G. Ragsdale
ontario institute for studies in education

A myth may have an empirical basis through chance occurrence. Some people would classify Aptitude-Treatment Interactions (ATIs) as such a myth. ATIs have great utility in describing, planning, and implementing instruction as well as their intuitive appeal. However, few disordinal interactions have been found, and those that have been found may apply to only a small portion of the population. In view of the scatter-gun approach that has had limited results, it is suggested that the ATI search be narrowed, with replications and estimates of effect sizes emphasized.

Les phénomènes de probabilité confèrent parfois aux mythes un fondement empirique. Certains classifieraient comme tel les études faites à partir de "Aptitude-Treatment-Interactions" (ATIs). Les ATIs sont fort utiles pour décrire, planifier et mettre en oeuvre les programmes d'enseignement; de plus, ils exercent un certain attrait auprès de ceux qui les utilisent. Cependant, on a trouvé bien peu d'interactions dysfonctionnelles et celles-ci ne s'appliqueraient qu'à une portion minime de la population. Considérant qu'une approche à cible multiple n'engendre que des résultats limités, on suggère ici de restreindre les champs de recherche en ATI tout en mettant l'accent sur la multiplication de ces expériences et en tenant compte des effets que peut avoir une population plus large sur les résultats.

Astrology and its adherents can be a source of frustration for the educational researcher. The frustration is increased because the millions who have faith in astrology are so much more numerous than the tiny band that has faith in educational research. It is particularly frustrating when empirical evidence is cited to prove the case for astrology.

The source of the evidence for astrology is a potent one, since a newspaper astrology column will produce 12 horoscopes (each containing several predictions) for each day of the year. If this output is multiplied by the hundreds of astrologers generating horoscopes, the number of amazingly accurate predictions that can be generated by chance alone is likely to be very impressive. Forer (1949) has shown, however, that personal validation is itself a fallacy. He produced a "personality sketch" which was accepted as accurate (for the individual) by 38 out of 39 students in an introductory psychology course.

More recently the concept of biorhythms has emerged as a predictive method similar to astrology. Briefly, the biorhythm theory assumes that

*This is a revised version of a paper presented at the Canadian Society for the Study of Education Annual Meeting, Saskatoon, Saskatchewan, June 6-8, 1979.

humans are affected by physical (23 day), emotional (28 day), and intellectual (31 day) cycles that begin (like a sine wave at zero degrees) at some neutral point at birth and then oscillate up and down to affect daily peformance. Articles describing this theory (in *Reader's Digest* and similar learned journals) often offer empirical evidence of the form: "The XYZ company found that 60% of their employee accidents occurred on the employee's 'critical' biorhythm days."

One question that might occur to the educational researcher if feelings of scientific outrage get too strong is, "How does one show that some effect does not exist?" We may suspect that those results showing an effect of biorhythms are merely due to chance fluctuations about a null effect, but proving our suspicions will probably be difficult.

One difficulty that we will encounter is the model of enquiry which we normally use. That is, we: (1) assume that there is no effect; (2) seek evidence to reject the assumption; and (3) if the evidence is found, label this as evidence to show the existence of an effect. All of our efforts are directed at rejecting the null hypothesis and the logic of our method dictates that we cannot accept the null hypothesis.

As an example of the difficulties involved in proving no effect, let us consider the following set of data. Most students applying to the Graduate Department of Educational Theory, University of Toronto, take the Miller Analogies Test (MAT). The test result card contains the date of testing, date of birth, and the raw score, in addition to some other information. Some 1,585 of these results were collected and biorhythm scores for the date of testing were calculated. In addition to the three biorhythm scores, day of each biorhythm cycle and astrological sign were also used as a basis for analysis of variance. In none of these analyses were there any results that approached statistical significance.

In fact one of the analyses, MAT score by astrological sign of birthdate, produced an F-ratio of only .76. However, it is difficult to use this information to produce an article that would appeal to any type of editor. One is more likely to wind up with a publishable article if one has faith in biorhythms (or astrology) and keeps sampling until the "true effect" is found.

FAITH VS. EMPIRICAL PROOF

Although Reynolds (1970) describes perpetuation of the myth that "women are capable of making finer visual color discriminations than men" and states that no evidence can be found to support the assertion, scientific myths usually require some empirical basis. There are at least two ways in which belief in a phenomenon can help to supply empirical evidence that the belief is justified. One is a belief which produces the effect directly, while the other is a belief which influences the search for effects. The more powerful of these ways is that in which belief acts

directly to produce the effect itself. This action of faith can be observed in a number of market situations. The following oversimplified version of the gold market is an example.

Very few purchases of gold bullion are made for the purpose of consuming the gold. It is unlikely that gold would be purchased if the purchaser expected it to decline in price, or even to remain stable, since essentially no-risk investments would provide a better return. Therefore we can assume that most purchasers buy gold because they expect the price to rise. This assumption is less than startling, but the point is that since the supply of gold is limited, an increase in the number of purchasers who believe the price of gold will rise usually results in a rise in the price of gold. Conversely, if more (potential) purchasers believe the price of gold will not rise, or fewer (potential) purchasers believe the price will rise, the price will probably fall. Thus there is a strong case for arguing that the price of gold has little to do with the value of gold and everything to do with the beliefs of potential purchasers. In short, if a prophecy is believed it will tend to come true.

The effect of faith is also important in other markets such as the stock and commodity exchanges. In the case of the commodity market, which has been described (particularly in the case of commodity futures) as a device to stabilize prices, it is interesting to note that the effect of faith on prices is probably destabilizing. In particular, if an "expert" predicts a rise in the price of coffee, an initial rise will confirm the expert's validity and lead to a larger rise. Similarly, any further predictions by this same expert should lead to larger effects, greater variance in prices, and less price stability due to the "positive feedback" effect (to use a cybernetic analogy) produced by confirmation of the prophet's judgement. Fortunately, there are often competing sources of information that tend to cancel each other's effectiveness.

The second way in which belief in a phenomenon can help to provide evidence of its existence is more applicable to educational research. This process, which many would say applies to biorhythms and astrology as well, is one in which a number of searchers (and researchers) are convinced that an effect does exist and set out to show that it does. (Note that although this process applies most directly to the question of existence of an effect, it can also enter into a dispute between two conflicting theoretical viewpoints if an expert can cause a majority of researchers to look for evidence supporting one of the points of view.)

This second process by which belief can appear to prove existence is perhaps best defined by an example. Suppose that some international cartel wished to convince people that drinking distilled water could increase the risk of cancer of the stomach. One strategy for the cartel would be to fund a large number of independent investigations of this hypothesis, trusting that the effects of Type I errors and our penchant for publishing significant results would bring about the desired result. The

results might also be affected by Rosenthal's (1978) finding that two-thirds of all recording errors favored the hypothesis being tested.

It is interesting to note that there is also an effect of "positive feedback" in the case of educational research fads. As a few positive findings are published in a new area of investigation, more researchers rush in to increase the flow of positive results. In the case of the previous example, it would mean that as the initial evidence of the link between dstilled water and stomach cancer became available, more researchers would begin to test this relationship.

The positive feedback also applies to the relative influence of educational research prophets. If a prophet is confirmed by a respectable number of positive results after one pronouncement, the flurry of activity after a further statement will be even more vigorous. Again, we are fortunate in having a number of competing leaders so that the influence of any one voice is attenuated by the calls from other leaders.

THE CASE FOR ATIs

One area of educational research which some might view as an instance of an empirically based myth is aptitude-treatment interaction (ATI). An interaction between treatment and aptitude occurs when the relative outcomes for the treatments are a function of the subject's initial aptitudes. Clark (1975) provides an example of a very real interaction in describing medical research into the effects of blood transfusions (blood type of transfusion interacts with blood type of patient). Cronbach (1957) described a similar type of interaction which he felt psychologists should be investigating as a result of his work in personnel psychology (Cronbach & Gleser, 1965). In this interaction, Treatment 1 produces the better result for one part of the aptitude range, while Treatment 2 is superior for another part of the range (see Figure 1). Lubin (1961) is credited with creating the name "disordinal interaction" for this type of effect where the lines of the plot actually cross.

The development of the ATI myth, if it is a myth, does not follow exactly the same pattern as those previously described. One important difference is that many more unsuccessful ATI investigations are reported because the presence of a significant main effect may be sufficient to achieve publication of the results. Bracht (1970) reported that in 90 studies which tested for 108 possible ATIs, there were only 5 which were significantly disordinal. It should be noted, however, that Bracht was using a more stringent definition of disordinal interaction than is usually the case. The criterion, developed by Bracht and Glass (1968), requires that differences between alternative treatments at two levels of a personological variable are both significantly non-zero and different in algebraic sign. Glass (1970) also commented on ATIs by saying, " 'There is no

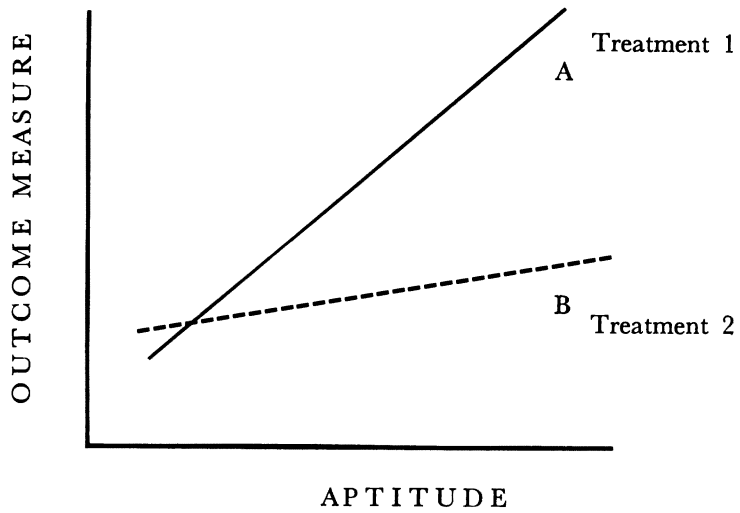

FIGURE 1

A Disordinal Interaction Between Aptitude and Treatments 1 and 2

evidence for an interaction of curriculum treatments and personological variables.' I don't know of another statement that has been confirmed so many times by so many people" (p. 210).

Although many negative ATI results are reported in the literature, the result is not particularly damaging to the credibility of the concept since ATI is a class of effects rather than one effect. The existence of ATI as a class of effects is the second major difference between it and the previous examples. As a result, negative results tend to be seen as disconfirming a particular instance of ATI, while positive results, even if they may be examples of Type I errors, can be seen as evidence of the existence of the overall concept.

Cronbach's original call for more ATI research did not have strong immediate effects in the area of instruction. In fact, according to one of the authors, the study by Ripple, Millman, and Glock (1969) may have been the first one specifically designed to test the trait-treatment interaction hypothesis. Nevertheless, after a slow start the area is now booming and as such is the main focus of the rest of this paper. Consequently, in all further discussion the use of "ATI" will mean interaction between aptitudes and methods of instruction.

Support for the existence of ATIs seems to be widespread throughout many areas of education. It is certainly intuitively appealing that there should be "different strokes for different folks" in instruction as well as in golf and tennis. This widespread appeal is indicated in the popularity

of what is generally called "individualized instruction." In practice, however, individualized instruction often allows for differences in rate of instruction and little else.

The emergence of the use of computers in education has served as a catalyst to the search for additional means of individualizing instruction. The early promotions for computer-assisted instruction (CAI) emphasized the notion that a computer could tailor instruction to the characteristics of the student. This is, of course, true, but before it can happen it is necessary that humans know how to perform this type of individualization. As more researchers rediscover the power of the computer, the dissatisfaction with this gap in educational knowledge will continue to mount.

One of the leaders in the individualized instruction field, Robert Glaser (1972), has called for the "new aptitudes," namely those that will interact with methods of instruction in the proper way. The approach which he recommended was "a conceptualization of individual difference variables in terms of the process constructs of contemporary theories of learning, development, and human performance" (p. 9).

In a later article, Glaser (1976) expanded his view to describe a psychology of instruction in which, although there was little mention of ATIs, there was emphasis on the analysis of cognitive processes that contribute to intelligence and aptitude-like performance and the implication (Glaser, Note 1) that performance analysis, individual differences, and learning should be subjected to analysis by process theories that can connect them.

Further definition of the area to be mapped out by ATI researchers was given by Salomon (1972). Salomon commented on the rapid increase in the number of studies in which ATIs are deliberately sought and offered three models to be used in generating ATI hypotheses. The models were remedial (to strengthen a student's weaknesses), compensatory (to instruct in spite of the weaknesses), and preferential (to cater to the student's strengths).

Finally, Cronbach (1975), Cronbach and Snow (1977), and Snow (1977) provide ATI researchers with ample encouragement, and some specific instructions, to continue the battle and provide the necessary information for the incorporation of individual differences into an instructional theory.

One of the important questions that arises is, "What ATIs are investigated?" The answer to this question is not simple because many types of ATIs are investigated. Some of the studies might be criticized by those who feel that the study has demonstrated the presence of an obvious interaction and the finding does not produce any new knowledge. For example, Kress and Gropper (1966) compared eighth grade students on programmed instruction materials presented (via television) at four fixed

rates. They found that when they compared those students who normally adopted (under self-pacing) a "slow" rate with those who normally adopted a "fast" rate on two treatments of fast and slow pacing, there was an interaction between external tempo (pacing) and characteristic work rate on retention scores. In other words, those students who normally adopted a slow rate did better when the pace was slow and those who normally adopted a fast pace did better when the pace was fast. The finding of this interaction (not classified as disordinal by the Bracht and Glass criterion) might be regarded by some as less than startling.

On the other hand, the search for surprising results may have become too enthusiastic. Shapiro (1975) comments on a general lack of theory in ATI studies and specifically mentions those studies that look for interactions between some treatments and a large battery of tests. She suggests that greater effort on a few tasks would be more fruitful. Cronbach and Snow (1977) also call for greater sophistication in the generation of hypotheses. The type of study that uses one set of materials with diagrams and looks for an interaction with spatial reasoning is given as an example of the low level of sophistication.

Cronbach and Snow (1977) have made a number of suggestions after analyzing the relationships between aptitudes (broadly defined) and instructional methods. In surveying the field, they found that many of the interactions that had been reported washed out with proper analysis. They also suggested that, although many researchers look to specific aptitudes for study, the general aptitudes offer better opportunities for useful results.

Although their prognosis for ATI research was generally positive, Cronbach and Snow warn that greater attention must be paid to higher-order interactions. Just as a significant second-order interaction can alter the interpretation of a significant main effect, so a third-order interaction can impinge on the interpretation of a second-order interaction. For example, Salomon (1972) cites Marshall (1969) as an example of the "preferential" model of ATI. However, the Marshall study, in addition to a significant main effect of reinforcement and an interaction between reinforcement and educational environment, also has a significant third-order interaction between reinforcement, educational environment, and sex. Thus the Marshall study can also serve as an example of how the meaning of an ATI can be altered by another ATI.

If higher-order interactions are common, then the usefulness of any particular interaction is diminished. Snow (1977) concludes that the prevalence of higher order interactions makes a general theory of instruction impossible to attain. He feels that a theory of instruction should be concerned only with "narrowly circumscribed local instructional situations, relatively small chunks of curriculum for relatively small segments of the educational population" (p. 12). Gelbach (1979) points out that

a general theory of instruction is impossible only if it must be prescriptive, and advocates instead "AT correcting," altering instruction to minimize the slope of Aptitude-Treatment regression while maintaining the mean level of achievement.

Finally, there is one further comment from Cronbach and Snow (1977) that seems to decrease the probability of any widespread use of ATI information. They state that aptitudes should not be regarded as stable, that aptitude develops as training proceeds. Therefore any assignment to a particular treatment should be reconsidered during the course of training. This shortens the terms of any benefits that are derived from a decision based on ATI information and decreases the overall value of the ATI approach.

ATTACKS ON ATI

Attacks on ATI research and theory have come from a number of different sources and only a few can be considered here.

Scandura (1978) has argued against the ATI approach primarily on theoretical grounds, presenting Structural Learning Theory as a more promising source of solutions to instructional problems (Scandura, 1978). Millman (1974) has taken an even more fundamental opposition to ATI work in saying that in general it is not a good idea to adapt teaching methods to student characteristics. One basis for this rejection is that "If we assume a normal distribution on the characteristic, relatively few students are located at the extremes where the biggest differences are expected when interactions are disordinal," (p. 214). Note however that Bracht (1970) and Cronbach and Snow (1977) have pointed out that cost considerations could shift the crossing point for a disordinal interaction (or turn an ordinal interaction into a disordinal interaction) and thereby change the rationale for the use of ATIs in instructional decision making.

Merrill (1975) takes a slightly different approach in arguing that ATIs are irrelevant. He challenges three assumptions, which he attributes to Cronbach and Snow (working from a pre-publication draft of their 1977 book), in making his case. First, Merrill asserts that dynamic aptitudes are more likely to produce useful results than studying stable aptitudes. Second, he states that varying strategies and tactics are more useful than fixed, repeatable procedures. Finally, he suggests that it is preferable for the individual to choose the environment, rather than have the environment adapt to the individual. One of his arguments for this last point is that if an instructional system presents all material in an optimal way, the learner will not be prepared for "real-world" experiences and will not know how to choose the best methods for learning in other situations.

SUMMARY AND CONCLUSIONS

The major points of the case against ATI research can be summarized quite briefly. Bracht (1970) has noted that few disordinal interactions have been found, and although Cronbach and Snow (1977) disagree with Bracht's analysis, they do agree with that conclusion. Millman (1974) states that the ATIs that are found will only apply to a small portion of the population because there are few cases at the extremes of the aptitude distribution. Cronbach and Snow (1977) and Snow (1977) point out that the use of a particular ATI may be restricted to a special setting because of the presence of higher-order interactions. Cronbach and Snow also indicate that an assignment based on an ATI may only be valid for a short period of time, since aptitudes are likely to change over the course of training. Finally, Merrill (1975) asserts that even if you do find ATIs, they are irrelevant to instructional decision making. In short, ATIs may be hard to find, apply to a few people in rare settings, be unstable over time, and be irrelevant to instruction — not a very promising basis for a theory, as Scandura (1978) has noted.

There are, however, arguments to be made in favor of ATIs, not the least of which is that individuals do have differences and that these are reflected in different needs. Probably the most compelling argument for the existence of ATIs is their great utility as an explanatory concept in describing, planning, and implementing instruction. There is also an intuitive appeal, but this appeal is not as strong as it could be. That is, if a student is high in some ability, it is reasonable to think of designing materials to capitalize on that strength, but less reasonable to think of designing materials for students who are low in that ability.

For example, suppose we had designed some materials that would capitalize on a student's quantitative ability. In conjunction with this we might also design other materials which could not just be low in quantitative demands, but would have to be high in something, such as demand for verbal skills. If we tested these materials on students classified as high or low in quantitative ability we would be losing a great deal of information. Our results would be obscured because, due to the correlation between quantitative and verbal skills, many "low quantitative" students would have poor results on the non-quantitative materials, and many "high quantitative" students would have good results on the non-quantitative materials.

Cronbach (1970) describes this problem in terms of choosing the correct test (aptitude) so that it will interact with the given treatments. However, unless one is dealing with a bipolar aptitude that has a strength at each end, it might be more logical to base the assignment to treatments on the relative strengths of two aptitudes. This type of inquiry may not lead us to a more fruitful model of instruction because the use of two aptitudes would lead to the familiar "unreliability of difference scores"

problem (see, for example, Ferguson, 1959, p. 285). It should illustrate, however, a continuing problem in the pursuit of useful ATIs, namely that the usefulnes of the aptitude being tapped by one set of instructional materials is diminished to the extent that it is positively correlated with the aptitude being tapped by the "alternate" materials. This implies that further ATI research should concentrate on instructional materials based on pairs of aptitudes that are not positively correlated or, even better, those that are negatively correlated, in order to maximize the probability of useful results. It is also possible that this suggestion is impractical, due to a shortage of aptitudes that are not positively correlated, and would only lead to further problems.

The confusion about ATIs is only symptomatic of a more general problem with our method of scientific inquiry, and any solution must address the more general problem. One way to begin is to heed the warning of Cronbach (1975) that "the time has come to exorcise the null hypothesis" (p. 124). This solution applies, as Cronbach originally intended, to all psychological research, not just to ATI studies, but the question remains, "Who will play the role of the exorcist?" (particularly if there is a film, "The Exorcist Meets the Null Hypothesis").

Since we desire to change human behavior and control of human behavior requires control of reinforcement, the journal editor seems to be a prime candidate for the role of the exorcist. If studies which present estimates of effect sizes are seen as being more publishable than hypothesis testing studies, the change in behavior should be accomplished with dispatch.

The task of convincing journal editors that it is preferable to have authors present estimates of effect sizes should not be difficult, since journal editors are extremely reasonable people.* The main thrust of the argument would be that a confidence interval (as an estimate of the size of an effect) gives much more information than does an hypothesis test. Not only is the hypothesis test information contained in a confidence interval, in that an interval including zero is not significant while one excluding zero is significant, but there is also additional information. For example, a confidence interval that covers the region from −.10 to +.10 conveys much more information than one which extends from −10.0 to +10.0 (assuming they are on the same scale), even though they are equal in terms of failure to reject the null hypothesis, because the first estimate indicates that the true effect is likely to be very small. The second estimate leaves a large effect, several points positive or negative, as a relatively likely event.

Therefore the major conclusion of this paper is that change in our method of scientific inquiry must be brought about through journal

*Editor's Note: The author, it will be recalled, is the former Editor of CJE/RCE, and therefore undoubtedly speaks the truth.

editors, while papers such as this one, which try to convince authors that the null hypothesis should be exorcised, are relatively useless.

REFERENCE NOTE

[1] Glaser, R. Personal communication, September 11, 1978.

REFERENCES

Bracht, G. H. Experimental factors related to aptitude-treatment interactions. *Review of Educational Research*, 1970, *40*, 627-645.

Bracht, G. H., & Glass, G. V. The external validity of experiments. *American Educational Research Journal*, 1968, *5*, 437-474.

Clark, R. E. Adapting aptitude-treatment interaction methodology to instructional media research. *AV Communication Review*, 1975, *23*, 133-137.

Cronbach, L. J. The two disciplines of scientific psychology. *American Psychologist*, 1957, *12*, 671-684.

Cronbach, L. J. *Essentials of psychological testing* (3rd ed.). New York: Harper and Row, 1970.

Cronbach, L. J. Beyond the two disciplines of scientific psychology. *American Psychologist*, 1975, *30*, 116-127.

Cronbach, L. J., & Gleser, G. C. *Psychological tests and personnel decisions* (2nd ed.). Urbana: University of Illinois Press, 1965.

Cronbach, L. J., & Snow, R. E. *Aptitudes and instructional methods: A handbook for research on interactions*. New York: Irvington, 1977.

Ferguson, G. A. *Statistical analysis in psychology and education*. New York: McGraw-Hill, 1959.

Forer, B. R. The fallacy of personal validation: A classroom demonstration of gullibility. *Journal of Abnormal and Social Psychology*, 1949, *44*, 118-123.

Gelbach, R. D. Individual differences: Implications for instructional theory, research, and innovation. *Educational Researcher*, 1979, *8*(4), 8-14.

Glaser, R. Individuals and learning: The new aptitudes. *Educational Researcher*, 1972, *1* (6), 5-12.

Glaser, R. Components of a psychology of instruction: Toward a science of design. *Review of Educational Research*, 1976, *46*, 1-24.

Glass, G. V. Discussion. In M. C. Wittrock & D. E. Wiley (Eds.), *The evaluation of instruction*. New York: Holt, Rinehart, and Winston, 1970.

Kress, G. C. Jr., & Gropper, G. L. A comparison of two strategies for individualizing fixed-pace programmed instruction. *American Educational Research Journal*, 1966, *3*, 273-280.

Lubin, A. The interpretation of significant interaction. *Educational and Psychological Measurement*, 1961, *21*, 807-817.

Marshall, H. H. Learning as a function of task interest, reinforcement, and social class variables. *Journal of Educational Psychology*, 1969, *60*, 133-137.

Merrill, M. D. Learner control: Beyond aptitude-treatment interaction. *AV Communication Review*, 1975, *23*, 217-226.

Millman, J. Instructional planning and management. In J. Blaney, I. Housego, & G. McIntosh (Eds.), *Program development in education*. Vancouver: University of British Columbia, 1974.

Reynolds, L. T. A note on the perpetuation of a "scientific" fiction. In L. T. Reynolds & J. M. Reynolds (Eds.), *The sociology of sociology*. New York: David McKay Co., 1970.

Ripple, R. E., Millman, J., & Glock, M. D. Learner characteristics and instructional mode: A search for disordinal interactions. *Journal of Educational Psychology*, 1969, *60*, 113-120.

Rosenthal, R. How often are our numbers wrong? *American Psychologist*, 1978, *33*, 1005-1008.

Salomon, G. Heuristic models for the generation of aptitude-treatment interaction hypotheses. *Review of Educational Research*, 1972, *42*, 327-343.

Scandura, J. M. Structural approach to instructional problems. *American Psychologist*, 1977, *32*, 33-53.

Scandura, J. M. Reconciliation of cognitive psychology and ATI research. *Educational Researcher*, 1978, *7*(3), 3.

Shapiro, K. R. An overview of problems encountered in aptitude-treatment interaction (ATI) research for instruction. *AV Communication Review*, 1975, *23*, 227-241.

Snow, R. E. Individual differences and instructional theory. *Educational Researcher*, 1977, *6* (10), 11-15.

Ronald G. Ragsdale is an Associate Professor in the Department of Measurement, Evaluation, and Computer Applications at The Ontario Institute for Studies in Education, Toronto, Ontario M5S 1V6.