

Journal of
The American Society
for
Church Growth

Volume 10
Spring 1999

Book Reviews

Natural Church Development

Reviewed by John Ellas and Flavil Yeakley

Christian A. Schwarz is the Director of the Institute of Natural Church Development located in Germany. He is the author of *Natural Church Development*.¹ In this book, Schwarz discusses some church growth principles that he says are supported by a massive research project that he started ten years ago. This research, according to Schwarz, involved 1,000 churches in 32 countries located in six continents. In each of these churches, the pastor and 30 members filled out questionnaires. The reported results identified eight quality characteristics related to quantitative church growth. Schwarz claims that this research provides the first really accurate way of distinguishing between church growth myths and universally applicable principles. In the opinion of these reviewers, however, this book is fatally flawed by the pseudo-scientific way the material is presented.

The scientific method requires some things that Schwarz simply did not do. The most obvious of these omissions is that Schwarz does not provide enough detail for other researchers to replicate his study and see if they find the same results. *Natural Church Development* was written at a popular level for non-

technical readers. Because of this, one would not expect the kind of detailed reporting found in scientific journals. But if a writer is going to base claims on the results of scientific research, that writer is obligated to provide enough information so that other researchers can replicate the study. That information might be presented in notes, in an appendix, or in a separate report. But at a minimum, a book that claims to be scientific should at least tell the readers how to obtain the more detailed information. Schwarzar does not do this.

When one of the reviewers ordered *Natural Church Development* from the American distributor, he asked where he could find the statistical data. He was told that Christoph Schalk had written a paper presenting all of the scientific data on methodology, statistics, etc. But he was told that this paper has not yet been translated into English. When he asked about purchasing the paper in German, he was told that it is not available in the United States—even in German. That is not the way scientific research is presented.

Another major omission is that significance levels are not reported. *Natural Church Development* has a large number of graphs that show the percentages of members who answer questions about quality characteristics in a particular way. These percentages are shown for four groups of churches.

- High quality, growing churches;
- High quality, declining churches;
- Low quality, growing churches; and,
- Low quality, declining churches.

What is omitted, however, is any information about the statistical significance of the observed differences among these four groups of churches. Some of the differences appear to be rather small. Readers need to know what the probability is that an observed relationship (a difference between percentages, a difference between means, a correlation between variables, etc.) was random and would not likely be found again in repeated samples. In social science research, the .05 level of significance is the minimal standard. The .05 level means that an observed relationship would occur by chance only one time out of 20. Schwarzar reports far more than 20 such relationships. If all of these were significant only at the minimal level, then at least one would

likely have occurred by chance and probably would not be found in repeated samples.

The .01 level means that the observed relationship would occur by chance only one time out of 100 and therefore one can be much more confident that repeated samples would find the same thing. The .001 level indicates a still higher level of confidence since the observed relationship would happen by chance only one time out of 1,000. Reporting significance levels takes up very little space. The reports could be put in notes or an appendix. They do not have to clutter up the text and distract the non-technical reader. But in a book that claims to be scientific, it simply is not acceptable to omit something as important as significance levels.

Another thing that is omitted in *Natural Church Development* are the correlation coefficients. There are many places in the book where Schwarzar claims to have found a "strong positive correlation" or a "strong negative correlation," but he never tells us how strong. This is an important omission. The square of the correlation coefficient is the amount of explained variance. It is not enough to know that a correlation is statistically significant and therefore would likely be found in repeated samples. It is also important to know how much of the variance on each of the measurements is explained by the correlation with the other. For example, suppose that you are reading a scientific report about the relationship between a quality characteristic called "evangelism" and quantitative church growth. Suppose that you find the following report of the relationship: $r = .2, p < .001$.

What that tells you is that there is a positive relationship between the "evangelism" score and numerical church growth (as one increases, the other increases). Furthermore, there is only one chance out of 1,000 that the observed relationship was random and would not likely be found again in repeated samples. But remember that explained variance is the square of the correlation coefficient and .2 times .2 is .04. Only four percent of the variance in numerical church growth scores is explained by the variance in "evangelism." A correlation that explains less than four percent of the variance is not usually considered to be of much practical significance, even if the correlation is statistically significant at the .001 level. The correlation coefficients should have been reported and this could have been done without distracting the non-technical readers. They could have been reported in footnotes, endnotes, or an appendix. They could have

been included in a separate publication, but only if the readers of *Natural Church Growth* were told how they could get this important information.

The pseudo-scientific nature of *Natural Church Development* is revealed by what is included as much as it is by what is omitted. On page 18, Schwarz writes, "This study developed into the most comprehensive research project of the causes of church growth ever undertaken." This study, however, can only be described as a correlational study and one cannot make causal inferences on the basis of correlations. It may be, as Schwarz suggests, that the eight quality characteristics cause numerical church growth. But it is also possible that in churches that are growing numerically, members are more likely to have positive perceptions about the eight quality characteristics. In that case, numerical church growth would be the cause and the good evaluations on the eight quality characteristics would be the effect. Another possibility is that both qualitative and quantitative growth are caused by some other variable. In the kind of study that Schwarz did, it is impossible to distinguish between cause and effect. When Schwarz claims to have identified the causes of church growth on the basis of a correlational study, that clearly puts this book in the pseudo-scientific category.

There is another place where Schwarz may have revealed more than he intended to reveal. On page 19, as Schwarz discussed the ten-year history of this research project, he wrote:

One of the most important criteria for our research project was a high scientific standard. Christoph Schalk, a social scientist and psychologist, agreed to coordinate the project and serve as its scientific advisor, after having identified several weaknesses in the testing procedures we had used until three years ago. He drafted a new questionnaire with rigorous standards for objectivity, reliability, and validity, and used approved methods from social science for the analysis of the data (emphasis added).

This project is actually the fifth stage in a series of research projects begun ten years ago in German-speaking Europe. Although the methodology may have left much to be desired in previous years, we gained an initial understanding of church growth principles upon which the subsequent studies were based (emphasis added).

What this plainly says is that for seven years this research project used a questionnaire that was not objective, reliable, or valid and methods of analysis that were not acceptable, but the *research project continued to base subsequent studies on this inadequate foundation!* Having a qualified social scientist as project coordinator and scientific advisor for the last three years could have helped if the conclusions had been based on the research done *after* the problems were corrected and if the essential data had been provided to the readers. Unfortunately, it was not done in that way.

There are several things that readers have a right to expect when an author makes the kind of claims that Schwarz makes in *Natural Church Development*. Schwarz says that he has developed a questionnaire that scientifically measures eight quality characteristics. Having made that claim, Schwarz must demonstrate the reliability and the validity of the eight scales.

When social scientists are doing the kind of research that Schwarz attempted, they typically begin by generating a large number of items for the questionnaire. Then they test these to see which ones are correlated with the criterion of numerical church growth. They use factor analysis to see which items go together. That is how they identify the relevant scales. Schwarz appears to have developed his eight quality characteristic scales *before* the factor analysis was done. Then seven years later he had Christoph Schalk conduct a factor analysis to demonstrate the validity of his eight scales. That is not the way to do social science research.

Item analysis is the next step in this kind of research. The way people respond to a good item will predict how they will respond to other items on that same scale. This is usually done by calculating prediction ratios. Sometimes the really good predictors are given more weight than items with lower prediction ratios. There is a newer method that can be used in a study as large as the one done by Schwarz. It is called "Item Response Theory" scoring. Authors should explain the kind of scoring they use. Schwarz does not.

When there is more than one scale, social scientists test items on all the scales. If an item shows a significant correlation with more than one of the scales, that item is rejected. The problem with such an item is that one cannot tell which scale it is measuring. When items are tested and put together to make up the various scales, social scientists have to make sure that the scales

are relatively independent. If they are not, it is not possible to tell what is being measured.

An author making the kind of claims that Schwarz makes must demonstrate internal consistency for each of the scales. This involves what is called "split half reliability." Each scale is divided into two parts. If the scale has internal consistency, there will be a high positive correlation between the X half and the Y half. Once internal consistency is established, the author must demonstrate stability over time. This involves what is called "test-retest reliability." The way a person evaluates a congregation's leadership, ministry, spirituality, structures, worship services, small groups, evangelism, and relationships on one day should be about the same as the way that person evaluates these eight quality characteristics one week later or one month later.

In addition to split half reliability and test-retest reliability, there is one other kind of reliability that sometimes must be demonstrated. It is called "alternate form reliability." The questionnaire used by Schwarz was translated into 18 languages. One cannot simply translate items word-for-word from one language to another. Items must be developed and tested within the context of each language. Responses to a questionnaire in one language must be shown to be consistent with responses to the same questionnaire in another language.

One of the factors that influences the reliability of a scale is the number of items on that scale. In general, the more items there are on a scale the more reliable that scale will be. The Myers-Briggs Type Indicator, for example, has 94 scored items for just four scales. Schwarz claims to measure eight quality characteristics with one questionnaire, but he does not tell his readers how many items there are on each scale. With Item Response Theory scoring, the reliability of a scale is not dependant on the number of items. But if Schwarz used IRT scoring, he does not say so. On page 19 of *Natural Church Development*, Schwarz claims that Christoph Schalk did all the appropriate studies, but he does not give his readers the results of any of these studies. That is not the way social science research is done.

If a questionnaire is not reliable, it cannot be valid. It is possible, however, for an instrument to be reliable and not be valid. A watch that is always five minutes slow is reliable (consistent), but it is not a valid measure of time. The point is that reliability is the first essential, but it is not enough. There must also be evidence of validity.

Construct validity is the important kind of validity that must be demonstrated for the eight scales that Schwarz developed.² The important thing is that authors making the kind of claims that Schwarz makes must be able to demonstrate that their scales really measure what they claim to measure. When a pastor and 30 members of a church give a church a very positive rating on a construct like "evangelism," how do we know that the church is really evangelistic? There must be some external evidence showing that a questionnaire really measures what it claims to measure. Schwarz does not even claim to have such evidence. For all of these reasons, these reviewers conclude that *Natural Church Development* is pseudo-scientific.

But the problem with the pseudo-scientific nature of this book is not the only reason for the opinion that it is fatally flawed. Another problem is the "straw man" approach of the author. Schwarz frequently explains how that his position is the middle ground between opposite extremes. That is a very natural tendency. All of us like to think that we have followed the counsel of God in Joshua 1:8 and have not turned aside to the right or to the left. There are people far to our right who are too conservative. Their consciences condemn things that we think are acceptable. There are people far to our left who are too liberal. Their consciences permit things that we think are not acceptable. These are very self-centered definitions of terms such as "liberal" and "conservative," but in practice that is the way we use them. Most of us, however, try not to do it in print.

Throughout *Natural Church Development*, Schwarz attacks people to his right and to his left. He gives the impression that these people are a part of the Church Growth Movement. The trouble is that he does not tell us who these people are. The two of us who are writing this review think that we know the Church Growth Movement rather well and we do not know whom Schwarz is talking about most of the time. Perhaps Schwarz is simply talking about how the Church Growth Movement is *perceived*. If so, his "straw man" strategy simply reinforces false perceptions. That is not a responsible strategy to employ.

It is our opinion that there is really nothing new in the eight quality characteristics that Schwarz discusses. There is, however, a danger in building a church growth strategy on nothing more than the perceptions of a pastor and 30 members about these eight characteristics. There are many other things that need to be considered.

One of the other things that needs to be considered is that Schwarz' organization is training individuals to conduct diagnostic evaluations for churches using a diagnostic tool (questionnaires for 30 members and the pastor) that is predicated on insufficient data gathering for a thorough and accurate evaluation. Some individuals will recognize this condition; others will not. Consequently, churches are offered a service that cannot meet expectations.

While essential qualities of a healthy church are not in dispute, the eight factors generated in the item analysis do not include local contextual, national contextual, and national institutional factors. Also, the eight items represent only a part of the possible local institutional factors that influence growth trends. Critical variables such as leadership, staffing, facilities, finances, size barriers, etc. are overlooked. In addition, personal perceptions alone are highly suspect for providing an accurate and thorough analysis. Ask members of any church, "What is your greatest strength?" and the vast majority will respond, "We are a friendly church." Ask a first time visitor to evaluate the same church and the majority will perceive it as a very unfriendly church. Both are partially correct. Members are friendly with one another, but not so friendly to guests.

A thorough and accurate evaluation is never based solely on soft data (such as perceptions gathered through questionnaires). An evaluation should include available hard data. These are areas that produce quantifiable results such as membership, attendance, staff ratios, group ratios, baptism ratios, involvement ratios, etc. Along with hard and soft data analysis, a thorough diagnostic should include an on-site visit, interviews, and direct observations of ministry activities. Only when soft data are supported by a broad based data gathering process can the data be trusted to closely reflect reality.

Motivated by the above concerns, I (John) decided to compare the evaluation results of a comprehensive diagnostic with the Natural Church Development survey for my home congregation. The results confirmed my concerns. Members' perceptions of congregational strengths were highly inaccurate in numerous categories. For example, need-oriented evangelism had the fourth highest score of 54. Unfortunately, while the congregation is growing and morale is high, there has been little evangelism emphasis, training, or results in the past five years. The congregation averages only 2.3 conversions per 100 members per year,

while the average for growing churches is 5.0 conversions per 100 members per year.

I realize that one example is not adequate to draw a final conclusion. But it does raise in my mind some serious doubts about the efficacy of the NCD survey as a diagnostic tool.

Conclusion

Christian Schwarz in *Natural Church Development* is claiming to have discovered universally applicable principles of church growth. On page 19 Schwarz says, "To my knowledge, our research provides the first worldwide scientifically verifiable answer to the question, 'What church growth principles are true, regardless of culture and theological persuasion?'"

Both of the reviewers have urged a greater concern for applying the scientific method to church growth studies. But to our knowledge, no researcher in the field of Christian ministry has ever claimed such a grandiose accomplishment as discovering universal principles. At the same time, Schwarz offers no scientific data to support his conclusions. Therefore, we consider this resource to be pseudo-science until the concerns presented earlier are adequately addressed.

In addition, the eight essential qualities of a healthy church have been identified in numerous other studies and are not in dispute. However, the efficacy of Schwarz's diagnostic tool to accurately evaluate the health of a local congregation is in serious doubt. Because of the magnitude of these concerns, we cannot recommend *Natural Church Development*.

Reviewers

Ellas, John W. Address: Center for Church Growth, PO box 691006, Houston, TX 77269-1006. Title: Director. John earned a B.S. from Northwestern State University of Louisiana in 1970, a M.S. from Georgia Southern College in 1972, a M.A. from Harding Graduate School of Religion in 1981, and a D.Min. from Fuller Theological Seminary. Dr. Ellas is an author, consultant, workshop leader, and publishes *Church Growth Magazine*.

Yeakley, Flavil. Address: Harding University, 120 Lambert Circle, Searcy, AR 72143. Title: Professor. Dr. Yeakley is a past president of the American Society for Church Growth and the 1995 recipient of the Donald A. McGavran Award. Flavil earned

a B.A. and a M.A. from the University of Houston and a Ph.D. from the University of Illinois in psychology and speech communication. A minister with the Churches of Christ since 1950, Dr. Yeakley is one of the leading researchers in the Church Growth Movement.

NOTES

1. The American edition was translated by Lynn McAdam, Lois Wollin, and Martin Wollin, with a forward by Robert E. Logan, and with the U.S.A. copyright by the American distributor, ChurchSmart Resources, 350 Randy Road, Carol Stream, Illinois 60188-1831 [128 pages, four color hardback, \$19.95 retail].

2. Some text books talk about "concurrent validity," "predictive validity," or "criterion-related validity." These are just different kinds of construct validity.