W. James Potter

Cultivation Theory and Research

A Methodological Critique

Devoted to Research in Journalism and Mass Communication
MISSION STATEMENT

Journalism Monographs was established, according to founding editor Bruce Westley, to serve the AEJMC membership and scholars and readers in adjacent fields by publishing original, scholarly works which are too long as articles and too short or too specialized for book form. One of the goals of the monograph series from the beginning has been to publish scholarly work from the entire field, whether the methodology was historical, legal, behavioral, or whatever.

In line with these original goals, the editor has adopted the following specific objectives:
1. To publish original, scholarly works which are too long as articles and too short for book form.
2. To publish research that makes a significant contribution to the field.
3. To publish manuscripts that will appeal to a wide audience of AEJMC members.
4. To cover all the various areas of scholarly interest in AEJMC regardless of research method.
5. To make the monographs as readable as possible by using effective layout, as well as relevant photographs, diagrams, tables, and charts.

TO CONTRIBUTORS

Submissions to Journalism Monographs should include an original and two copies, double-spaced throughout (including extracts, references, and notes). Preferred length is 30-60 pages. The names of authors should appear only on the title page, and other identifying material should not appear in the manuscript.

Authors must use one of these three style guides: The Chicago Manual of Style, published by the University of Chicago Press; the Publication Manual of the American Psychological Association; or A Uniform System of Citation, published by the Harvard Law Review Association. The style chosen should be the one most appropriate to the subject matter of the manuscript submitted. Footnotes and references should be assembled at the end of the manuscript. Tables and figures should be on separate pages, not in the text.

Journalism Monographs is produced on the Macintosh computer using Pagemaker software. Once a manuscript is accepted, the author must provide the editor with the manuscript on a computer floppy disk. This can be either a Macintosh disk or an IBM or IBM-compatible disk containing a text-only file.

The editor is particularly interested in including more illustrative material in Journalism Monographs. Submissions of manuscripts using pertinent and high-quality photographs or well-done and effective charts, diagrams, or maps are welcomed.

Authors are expected to be candid with the editor in matters pertaining to the origins and previous appearances of manuscripts. Manuscripts under consideration elsewhere should not be submitted to Journalism Monographs. It is also policy not to publish a long version of a study published in a shorter version elsewhere.

Manuscripts should be sent to John Soloski, editor-elect, Journalism Monographs, School of Journalism and Mass Communication, University of Iowa, Iowa City, IA 52242.
The central proposition of cultivation theory is: Television viewers who say they are exposed to greater amounts of television are predicted to be more likely (compared to viewers who say they are exposed to lesser amounts) to exhibit perceptions and beliefs that reflect the television world messages. Researchers who have provided tests of this proposition have relied on two methods. Content analysis has been used to determine the frequency of certain messages in the television world. Then survey methods were used to ask viewers about their perceptions and beliefs concerning the real world.

Cultivation theory forces researchers to confront three important methodological questions: (1) How should television exposure be measured? (2) How should cultivated perceptions be measured? and (3) What is the appropriate test for the relationship between exposure and perceptions? The beginning point for answering each of these questions is to focus on the conceptualizations in the theory. This has been helpful up to a point (for a conceptual critique of the theory, see Bryant, 1986, and Potter, 1993). Most researchers, however, have seemed to rely more on the measurement and analysis practices in the empirical literature. While this literature has been very helpful in demonstrating that there is a cultivation effect, this empirical work also contains some troubling methodological problems.

Methodological problems are always expected when researchers open up new areas of study, especially with a theory that breaks as much new ground as cultivation theory did. Early empirical work must necessarily be regarded as exploratory. But as the literature grows, it should also mature. Critics are needed to influence thinking and practices constructively, and there have been some insightful criticisms of the early literature (for example, see Hawkins & Pingree, 1981, 1982; Hirsch, 1980, 1981a, 1981b). Many cultivation researchers, however, have been very slow in responding to this criticism, preferring instead to continue with faulty practices without confronting the criticism and providing a justification for their methodological decisions. In addition, the last decade has produced other findings that indicate there is a need to change the way cultivation research is conducted, especially with the tasks of operationalization, data gathering, and data analysis. Because of these reasons, it is important to present a unified picture.
of the methodological practices throughout the entire body of research on cultivation theory and to do so from a critical point of view. This monograph attempts to serve that function.

This monograph is organized into five sections. First, the measurement of television exposure is examined. Second, the measurement of cultivated perceptions is investigated. Third, the various methods of testing for the relationship are illuminated. Within each of these three sections, research practices are first described, then they are critically analyzed to show the importance of acknowledging the early critiques as well as subsequent findings in designing cultivation research. The next section deals with the issue of sampling. The final section of the manuscript presents five recommendations for future tests of cultivation theory.

**The Measurement of Exposure**

The early writings about cultivation contain no theoretical calculus to guide the operationalization of viewing. The creators of the theory, however, have conducted a series of studies which show a consistent operationalization of exposure as the number of hours of total television viewed on a typical day (for example, Gerbner, Gross, Jackson-Beeck, Jeffries-Fox, & Signorielli, 1978).

Over time, the literature on cultivation has come to display a great variety of approaches in terms of measurement (how the assessments are made), scaling (how responses from individual questions are combined into an index of television exposure), and data distributions (the treatment of the variable in the analyses).

**Measurement of Television Exposure**

The variable of television exposure has always been measured as a self-report item on a questionnaire; direct observation has never been used. Also, it is rare to see studies using repeated measurements of the same viewers over time (Jeffries-Fox & Signorielli, 1979; Potter, 1991c). It is more prevalent to see studies using a series of cross sectional measurements of different samples rather than a repeated measuring of the television viewing of the same respondents (for example Gerbner, Gross, Jackson-Beeck, Jeffries-Fox, & Signorielli, 1977; Gerbner, Gross, Signorielli, Morgan, & Jackson-Beeck, 1979; Gerbner, Gross, Morgan & Signorielli, 1980; Morgan, 1986; Signorielli, 1990).

Within these constraints of cross-sectional self-reported data, there has been a variety of questions used to generate those data. These questions or items can be categorized into five groups. Only the first of these (global assessment) is a faithful operationalization of exposure as originally conceptualized in the theory. The other four represent either useful extensions of the theory or faulty operationalizations, depending on one’s point of view. **Global assessment:** Gerbner and his colleagues typically operationalized exposure by asking people to assess total viewing of television regardless of program type or source. In most of their reports of cultivation analysis, Gerbner and his colleagues have analyzed data gathered by the National Opinion Research Corporation (NORC), which used the following TV exposure question: “How many hours of television do you watch on an average day?” This question has been used often by Gerbner’s colleagues and former students (Gross & Jeffries-Fox, 1978; Morgan, 1983, 1984, 1986; Signorielli, 1990) as well as others (Carveth & Alexander, 1985; O’Keefe, 1984; Tan, 1982; Volgy & Schwarz, 1980). Also, there are slight variations on this question, such as “How many evenings per week do you watch TV at least one hour?” (Fox & Philliben, 1978).

**Exposure to genres:** Potter (1986) presented respondents with a list of 11 television genres (action/adventure, situation comedy, afternoon soap operas, prime time soaps, news/documentaries, movies, sports, cartoons, music, game shows, and talk shows) and asked them to estimate the number of hours they viewed on a weekly basis within each program type.

**Exposure to particular shows:** Some studies present their respondents with a long list of programs and ask them to check the programs they regularly view (Elliott & Slater, 1986; Ogles & Sparks, 1989; Slater & Elliott, 1982); some ask respondents to check the shows they have viewed in the past seven-day period (Doob & Macdonald, 1979; Weaver & Wakshlag, 1986; Wober, 1978); some ask respondents to check whether they watch the show every time, a lot or once in a while (Carlson, 1983; Eron, Huesmann, Brice, Fisher, & Mermelstein, 1983; Huesmann, Lagerspetz, & Eron, 1984; Volgy & Schwarz, 1980); one asked respondents to use a Likert scale (1 = “never” to 5 = “usually”) (Rubin, Perse, & Taylor, 1988); and one asked respondents to specify whether they viewed the show every week, most weeks, some weeks, or never (Reeves, 1978). Weaver and Wakshlag (1986) administered a prior week prime time television viewing diary which listed 65 programs, and respondents checked all the programs they viewed in the previous seven-day period.

The most complete form of measuring exposure was used by Hawkins and Pingree (1980, 1981), who asked their respondents to fill out a viewing diary. All programs were listed, and respondents wrote the number of minutes they viewed each program during the measurement period.

A few studies were only interested in exposure to particular shows within a genre. For example, Carlson (1983) presented a list of 16 selected crime shows and had respondents use a 4-point scale to respond (“never” to “almost always”). Carveth and Alexander (1985) focused on the genre of soap operas and presented viewers with a list of 13 network soap operas where respondents indicated the number of episodes they watched in a typical week.
Long term exposure: Carveth and Alexander (1985), in addition to the measure described above, also asked respondents when they began viewing soaps, and this was converted into a measure of years of viewing.

Exposure through attention: Rouner (1984) tested cultivation with a unique measure of exposure. She developed a model based on attention to television rather than hours viewed. Specifically, she asked respondents how much attention (“close attention” to “no attention”) was paid to the “personal qualities of the characters, the characters’ appearance, values and morals displayed by the characters, and the story line” (p. 171).

Critique of Exposure Measures

The most troubling aspect of measurement on the exposure variable is the ambiguity introduced by some researchers who used a Likert scale ranging from “almost always” to “practically never” view (Allen & Hatchett, 1986; Carlson, 1983; Tan, 1982; Volgy & Schwarz, 1980). It is puzzling that researchers did not simply ask respondents how many hours of television they watch on an average day or during an average week. This would be the simplest, most direct measure, and the resulting data would provide the greatest number of options in the analysis. When researchers choose to deviate from this most defensible form of measurement, they need to provide a clear rationale for doing so. The Likert-type of measure is less precise than asking about hours viewed. Perhaps there is some reason why these researchers felt that a less precise measure was warranted, but they gave no rationale for use of the weaker measure.

Scaling

The theory contains no prescriptions for scaling or analysis, but again operational practices have developed to guide research. Not all cultivation researchers conform to the practices used by the theoreticians, however. In some studies, the single measure of viewing per day is used as the exposure measure, but in other studies several measures are aggregated in some way to construct a viewing scale. There are three forms of aggregation.

Summative scales: Weaver and Wakshlag (1986) computed two indexes: non-crime related viewing and crime related viewing. For every hour-long crime show a respondent checked in his/her seven-day diary, 60 minutes were added to the crime viewing index. A similar procedure was used to translate check marks on non-crime shows into index minutes. The authors report that reliability was marginal for crime related viewing (KR-20 = .64) and non-crime related viewing (KR-20 = .69). But the Kuder-Richardson test is a split half method of testing internal consistency of a scale. It splits the items on the scale into two sets and examines the degree to which measures on one half of the items are related to measures on the other half of items. It does not seem that the results of Kuder-Richardson would address the issue of reliability here, because there is no reason to expect someone who watches one crime show to watch all other crime shows. So a low KR-20 coefficient is less an indication of the quality of the scale than it is an indication that viewers have differential viewing preferences.

Hawkins and Pingree (1980, 1981) summed the minutes recorded in respondents’ diaries to construct six program type exposure scales: news, situation comedies, crime adventures, dramas, game shows, and cartoons. They also had respondents use a 4-point scale (“never” to “often”) to indicate how often they viewed programs in these seven genres. They found that “except for cartoons (r = .46), correlations between the two forms of measurement were quite low and inconsistent” (p. 205). Given the choice of the two measures, they concluded “while the two methods of measurement do not relate well to each other, viewing of the content types measured by the TV diary is better related to the cultivation dependent variables, suggesting that the diary is also a better measure of long-term viewing habits than the averaging measures” (p. 205).

Several researchers have computed summative scales where the elements summed were not hours of exposure. For example, Carlson (1983) summed the responses (1 = “never watched”; 4 = “almost always watch”) to 16 crime shows to construct a Crime Show Viewing scale that had a range of 16 to 64. Volgy and Schwarz (1980) used responses to a list of 25 prime-time programs (how regularly viewed?) and computed exposure scales for medical and minority show viewing.

Multiplicative scales: Some researchers have computed a weekly viewing figure from daily viewing data by multiplying the daily viewing response by 7. For example, Carveth and Alexander (1985) used their respondents’ hours of viewing per day and multiplied by 7 to get the weekly hour figure. This scale is based on the assumption that viewing is the same all seven days of the week. Ogles and Sparks (1989) multiplied the weekday viewing measure by 5 and the weekend day viewing measure by 2 and then summed the two products to get a weekly figure.

Average exposure: Another type of scale is the average. For example, in two studies, exposure was operationalized as the average between two measures: amount of television watched yesterday (a weekday) and the amount of television that usually watched on an average weekday (Perse, Ferguson, & McLeod, 1994; Rubin, Perse & Taylor, 1988).

In another study, exposure was operationalized as the sum of hours of TV viewed on a typical weekday and a typical weekend day (Hawkins, Pingree, & Adler, 1987). This type of scale is merely an operational device that appears to be built on a feeling that a single item is not a good indicator of viewing, nor is any item by itself. This scale has no conceptual meaning, that is, it does not reflect daily or weekly viewing. It can also be misleading. For example, if person A watches 4 hours on a typical weekday and 6 hours on a typical weekend day, his score on this scale is 5. If person B watches 2 hours on a typical weekday and 8 hours on a typical weekend day, his score on the scale is 6. The scale indicates that B is a heavier TV viewer despite the fact that...
A has much higher weekly viewing and also watches more than B on five out of seven days of the week.

**TRANSFORMING EXPOSURE DISTRIBUTIONS**

In all of the research on cultivation, the exposure measure provides a continuous distribution of viewing, almost always expressed in terms of hours. Despite no guidelines from the theory, an operational practice of segmenting the continuous distributions of viewing measures began early on in cultivation research. Gerbner and his associates established a convention of constructing a categorical variable for exposure. This is curious because the measurement of viewing exposure yields a continuous distribution, but Gerbner and his colleagues typically divide the continuous distribution into discrete categories. The cut points vary from study to study, however. For example, Morgan (1986) trichotomized viewers into light (1 hour or less), medium (2 or 3 hours), and heavy (4 or more hours). Signorielli (1990) truncated her continuous data on exposure to categories as follows: light viewers (under 2 hours per day), medium (2 to 4 hours), and heavy (over 4 hours each day) so that the three groups were approximately equal. Volgy and Schwarz (1980) took their continuous data and categorized respondents as light (one or fewer hours per day), moderate (more than 1 but less than 3), and heavy (3 or more). Ogles and Sparks (1989) constructed four groups from weekly viewing totals: low (0-2 hours per week), moderate (3-9 hours), heavy (10-19 hours), and excessive (20-55 hours).

Sometimes the cut points vary between analyses within a single article. For example, Gerbner et al. (1978) define light viewers as those respondents viewing 2 hours or less per day, medium viewers as those viewing 2 to 6 hours, and heavy viewers as those watching 6 hours or more. In another analysis published in the same article, they offer yet another operationalization as follows. In answer to the question “How often do you watch evening entertainment programs about crime and police?” light viewers were those respondents who said “rarely or never,” medium viewers were those who said “sometimes,” and heavy viewers were those who said “frequently.”

Also, there is not a common practice for determining how many viewing groups there should be. In various studies, respondents are placed in groups of high, medium, and low users (Gerbner, Gross, Morgan & Signorielli, 1982; Volgy & Schwarz, 1980), into two groups of high and low (Gerbner et al., 1979), or into four (Ogles & Sparks, 1989).

Fortunately, not all cultivation studies divide their continuous distributions of exposure into categories. There are many examples where hours of viewing are measured and analyzed as a continuous variable (Carveth & Alexander, 1985; Christiansen, 1979; Gross & Jeffries-Fox, 1978; Hawkins & Pingree, 1980, 1981a, 1981b; Hawkins, Pingree, & Adler, 1987; Morgan, 1983, 1984; Perse, 1986; Perse, Ferguson, & McLeod, 1994; Potter, 1986, 1988; Rubin, Perse, & Taylor, 1988; Tan, 1982). When the continuous distribution of this variable is preserved, it is defensible to use parametric statistics, such as Pearson correlation, partial correlation, and multiple regression.

With the measurement of television exposure, it is a serious problem to group viewers into exposure categories by using arbitrary and inconsistently applied cut points. If the relationship between exposure and cultivation was linear, then this would be a minor problem (perhaps leading to difficulty in synthesizing across studies) at most. But the relationship does not seem to be linear (see section below titled “Nature of the Relationship”), and this means that the analytical tests of differences and associations will be sensitive to where the cut points are drawn. For example, if the cultivation relationship is not linear, then the altering of the arbitrary cut points can change the results. The top part of Figure 1 shows that the light viewers have a lower cultivation score than medium viewers who have a lower score than do heavy viewers; also, the trend within each of three viewing groups is generally positive but weak. The lower part of Figure 1 shows the same relationship curve, but the cut points are drawn in different places. In this lower chart, the mean cultivation score of the medium and heavy groups could be the same; also the positive trend in the light and medium viewing groups is in contrast with the negative trend in the heavy group.

This practice of using arbitrary cut points and changing them across studies was criticized at an early date by Hirsch (1980), who observed “major inconsistencies in classifying light and heavy viewers across samples.” In reply to this criticism Gerbner et al. (1981) said, “We have never implied nor argued that the terms ‘light’ and ‘heavy’ viewer are anything but relative, determined by the distribution of responses in any given sample.” They continued, “Any attempt to specify ‘absolute’ levels of heavy viewing or absolute proportions of the sample is doomed to failure if these standards are applied to samples of different ages” (Gerbner, Gross, Morgan, & Signorielli, 1981, p. 46). And a decade later, the theorists are providing the same defense, “What is important is that there are basic differences in viewing levels, not the actual or specific amount of viewing” (Morgan & Signorielli, 1990, p. 20). But this misses the point about why the practice of changing cut points is faulty. The criticism is not calling for standard cut points. The criticism is that the moving of cut points can have the effect of changing the results, so researchers should be careful to report their results using a variety of alternative cut points, or better yet, simply show a plot of the relationship. This alternative practice of displaying graphic plots will help the research take a significant step towards focusing on the shape of the relationship rather than directing attention toward the weak strength of an assumed linear association.
If the cultivation relationship were linear, then the practice of transforming a continuous distribution into a categorical one would not be a critical flaw. But it would still be a curious practice. Why segment a continuous distribution? Transforming a distribution of a wide range of values greatly reduces the variance and limits the variety of statistical procedures that can be used. Perhaps the researchers have some outliers in the viewing distribution and fear that the influence of these outliers might have a detrimental effect on computing the relationship. But if this is the case, there are ways of “normalizing” the viewing distribution. If there is a reason for categorizing the distribution (and there may well be), the researchers should share that with their readers.

If, instead, the cultivation relationship is non-linear, a finding already documented in the literature, then transforming a continuous distribution into a categorical one serves to mask the relationship and can even lead to spurious results. If researchers insist on segmenting their distributions, they must provide a very strong rationale for the selection of the particular set of cut points. Furthermore, they should demonstrate that the use of another set of cut points would not alter the results.

**Cultivation Indicators**

**Measurement**

There has been a wide variety of measurements for cultivation indicators. Some items require respondents to offer an estimate, some ask for a perception, others focus on a belief, attitude, or value. These can be categorized according to what has been referred to as the “order” of the measure (Hawkins & Pingree, 1982). Estimates of some real-world parameter have been called first-order measures. In contrast, second-order measures are more global reactions to the real world such as perceptions, attitudes, feelings, and values.

**Estimates:** There has been a wide range of topics covered, including estimates of rates of crime and violence (Carveth & Alexander, 1985; Doob & Macdonald, 1979; Gerbner et al., 1977; Hawkins & Pingree, 1981a; Ogles & Sparks, 1989; Potter, 1991a, 1991b, 1991c); estimates about personal victimization through crime (Gerbner & Gross, 1976; Hawkins & Pingree, 1980; Morgan, 1983; Ogles & Sparks, 1989; Weaver & Wakshlag, 1986); estimates about risks in life, such as being hit by lightning, flooding, and terrorist bomb attacks (Gunter & Wober, 1983); estimates about proportion of people employed in law enforcement (Gerbner & Gross, 1976); estimates of numbers of professional women (Carveth & Alexander, 1985); estimates of rates of divorce (Carveth & Alexander, 1985; Potter, 1991b); estimates of affluence (Fox & Philliber, 1978; Potter, 1991b); and estimates about the aged (Gerbner et al., 1980).

**Perceptions:** There is also a wide range of topics focusing on how people see the world including perceptions about a mean world (Bryant, Carveth, & Brown, 1981; Morgan, 1986; Rouner, 1984; Signorielli, 1990; Wober, 1978; Zillmann & Wakshlag, 1985); perceptions about doctors (Volgy & Schwarz, 1980); perceptions about traditional sex roles (Volgy & Schwarz, 1980); perceptions about sexism (Gross & Jeffries-Fox, 1978); and perceptions about American stereotypes (Tan, 1982).

**Attitudes:** There have been studies of socio-political attitudes, such as approval of police brutality and bias against civil liberties (Carlson, 1983); attitudes toward blacks, personal conduct, communism, free speech, federal spending, and taxes (Gerbner et al., 1982); perceptions about black groups, black self-esteem, black group identification/mainstream, and the black separatist perspective (Allen & Hatchett, 1986); concern with racial problems (Volgy & Schwarz, 1980); attitudes about abortion, free speech restrictions...
After choosing a topic, researchers need to be very precise in the wording of their items.

They seem to imply that it is too hard for respondents to come up with estimates on their own. But if they believe this, then how can they justify that those estimates exist in the viewers' minds? If those estimates do not exist, but must be created by the measurement situation, then how valid can the data be when they are so artificially generated?

After choosing a topic, researchers need to be very precise in the wording of their items. A seemingly subtle change in the wording of a item can, in effect, change the topic, as has been demonstrated by Ogles and Sparks (1994). They presented their subjects with a list of 16 crime situations, such as being beaten up by a stranger, having your car stolen, and being threatened with a knife, club, or gun. They asked their subjects to rate each of these situations on a 1 to 10 scale in terms of “how afraid you are about becoming the victim of each type of crime in everyday life” (p. 54). Then they asked the subjects to rate each of these situations on a 1 to 10 scale “according to what you think your own chances are of being a victim of that crime” (p. 54). While these two sets of measures appear very similar, the first is a measure of fear of victimization and the second is a measure of probability of victimization. The resulting data indicated that the two sets were measuring two very different concepts. “On average, almost 80% of the variance in the fear-of-victimization measures was left unshared by the probability-of-victimization measures” (p. 59). Also, the former set was found to provide stronger evidence of cultivation on most of the items.

There is another problem with the items. There is a serious unmet concern about justifying what should be the television world answer. This is not a serious problem with first order measures. The television world answer is relatively easy to determine because it can be based on the number of occurrences of certain acts which are documented through content analyses. Once we have this count, we look up the real world parameter in the U.S. Census Report and if the two figures are very different, we have a good cultivation indicator. There is a major problem, however, with providing a priori justification for what the television world answer should be when using second order (or belief system) measures. Because second order measures require the respondents to infer patterns of behavior, how can cultivation researchers be confident that their designated television world answer has taken into account all the factors in that inferential process? This is a criticism leveled by Hughes (1980) and Newcomb (1978) who argue that the television world answer may be impossible to determine, because viewers take into account so many contextual elements of portrayals (not just the frequency) when they infer meaning from television programming. It can often be the case that viewers believe one answer to be the television world answer when Gerbner has designated another answer as the television world answer, so the results that were reported to support cultivation could easily be interpreted to be a refutation of the theory.

This problem, which is troublesome on the conceptual level, is even more serious on the operational level where practices exhibit two types of

Cultivation Theory and Research
flaws: non-justification and unwarranted justification. The non-justification is apparent in studies that provide no rationale for selecting a TV world answer. This problem of omission appears to be less serious than the problem of unwarranted justification from the analysis of televised content (see Newcomb, 1978, for a conceptual treatment of this problem). For example, Gerbner’s justification that the TV world is a mean and violent one may be unwarranted when viewed from a wider perspective that includes the context of acts in addition to their frequency. Context is very important to viewers. Gerbner acknowledges this point in his 1969 initial set of criteria, but his team largely ignores context when generating their annual counts of violence on television. Those content analyses focus on frequencies of violent acts. It is too simple, however, to infer that the viewers learn that the real world is mean and violent from television simply because there is a high frequency of violence there. Viewers could be looking at the context of those portrayals and draw their own conclusions that the world is peaceful, because criminals are always quickly removed from society either by being caught or killed.

Currently the theory is prescriptive in its requirement for an a priori determination of which answer is real world and which is television world. Instead, it might be more useful if the theory were more oriented toward generating descriptive evidence that identified what answers were most prevalent among the heaviest viewers. But in order to see such patterns, if they exist, a continuous distribution of cultivation indicators would be necessary. For example, it would seem to be far better to ask respondents to provide an estimate of the percentage of adults who are divorced (thrust producing a continuous distribution of responses from 0 to 100) than to ask them to choose between two numbers. Perhaps, many respondents feel frustrated when they are asked to provide an estimate when so many levels of response are available to them. It has been demonstrated that people have a hard time estimating mundane behaviors such as numbers of times they change channels (Ferguson, 1994); perhaps they would also have a hard time “figuring out” an estimate about their likelihood of being involved in a crime. But it is not clear why respondents would feel more confident in their answers if they were presented with a closed form answer choice (two inaccurate answers) than if they were presented with an open-ended answer format. This is one area where more research is needed. Psychometricians need to establish the comparative validity and reliability of responses to different types of answer choices to cultivation questions. But until this issue is settled, it would seem better to employ the open-ended answer choices that would generate more variance in responses.

Scaling

If a study uses only one item to measure respondents’ cultivation, then there is no possibility for scaling. For example, Morgan (1983) used one item: “During any given week, what are your chances of being involved in some kind of violence? about 1 in 10/ about 1 in 100?” But most studies use more than one item, and this then raises the issue of scaling. Does the researcher choose to treat each item separately in the analysis? Or does the researcher aggregate the data from multiple items, thus creating an index or scale? Studies are presented below in a three-part classification as follows: no scaling, good scaling, and questionable scaling.

No scaling: Several studies used the data from their items in the analysis without attempting to aggregate data, such as adding responses from several items together to form a summative scale. For example, Reimer and Rosengren (1990) looked at values of materialism and postmaterialism by using 12 items where respondents react to items using a 7-point Likert agree/disagree response choice. They computed their associations on all 12 items instead of summing the 12 to form a single index.

Carveth and Alexander (1985) asked respondents to estimate the number of females out of 100 in the U.S. population who were doctors, lawyers, housewives, divorced, and mothers of illegitimate children. They also asked for estimates of the number of men (out of 100) who were doctors, lawyers, policemen, divorced, and fathers of illegitimate children. Finally they asked for an estimate of how many marriages out of 100 end in divorce and how many people out of 100 have committed a serious crime. No scaling was done — all items were analyzed separately. Ogles and Sparks (1989) report the results of 8 individual items in one study and 16 items individually in another.

In contrast to their earlier work, Gerbner et al. (1978) began a practice of not using scales, but dealing instead with items individually. This seems odd, because they were using a three-item anomic scale based on the work of Srole (1956), a scale that had often been tested and validated and then included on the 1977 NORC General Social Survey. Without providing a reason, however, they chose to analyze the items individually and not to aggregate them into a scale as had been done before.

Good scaling: In order for researchers to demonstrate that their scales are good, they need to show that the individual items in the scale are internally consistent. . . .
Several studies do not make the case for good scaling.

Items loaded on one factor, and the alpha of the resulting summative scale was .85. Therefore we can conclude that these seven items are a consistent measure of an underlying factor of perceptions of affluence.

Morgan (1984) used a battery of 26 items asking respondents to describe their lives—boring, depressing, and so forth. The items were put into scales of great, calm, intense, and lousy and each had an alpha ranging from .70 to .77. In another study, he developed 12 scales: political self-designation, approval of abortion, free speech restrictions (anti-leftist and anti-rightist), racism, legalization of marijuana, reactions to federal spending, sexual tolerance, sexist beliefs, fear of local areas, perceptions of a mean world, and anomie. The alpha coefficients ranged from .90 to .53 with a median of .68 (Morgan, 1986).

Rubin, Perse, and Taylor (1988) report acceptable alphas on four of their scales, but they use all five in their analyses. The scales and alphas are faith in others, .91; life control, .75; interpersonal connection, .76; political efficacy, .80; and safety, .57.

Questionable scaling: There are three reasons for judging scales as questionable: (1) Internal consistency of the scales is not described or only partially described, (2) internal consistency data are presented but the coefficients are too low to conclude that the scales are of high quality, and (3) questionable psychometric procedures are used.

Several studies do not make the case for good scaling, leaving the reader with no information to judge whether the scales are internally consistent. For example, Allen and Hatchett (1986) constructed four social effects scales (black group perception, self-esteem, black group identification/mainstream, and black separatist perspective), but they report no reliability tests. Gross and Jeffries-Fox (1978) asked children to react (“true” or “false”) to five statements such as “women have less chance than men to get the education for top jobs,” and “our society discriminates against women” (p. 263). They constructed a scale from these items but how this was done was unreported, and they report no reliability data. Weaver and Wakshlag (1986) asked respondents to respond to 14 crime-related judgments which were a measure of the likelihood of victimization in typical and hypothetical situations. The response scale ranged from “not at all” (0) to “extremely” (10). A factor analysis with oblique rotation yielded three factors. Three scales (situational anxiety, environmental anxiety, and personal anxiety) were constructed using factor-scores as weightings, but no reliability data are presented. Hawkins and Pingree (1980) constructed a mean world index from viewers’ reactions (5-point Likert scale) to six items such as “If they got the chance, most people would try to cheat me” (p. 203). They performed a factor analysis and found only three items to load on a mean world concept, so only these were summed to form the index, but no reliability figures were given. Also, Rouner (1984) had her respondents use a 5-point Likert scale to react to three mean world questions. After running a LISREL analysis, she said that the items were found to be reliable, but she does not present any goodness of fit results so the reader can judge the degree of internal consistency among those items.

A second problem with questionable scaling occurs when researchers run tests of reliability, but the results of those tests do not support a case for a quality scale. For example, Carlson (1983) had 12 items measuring support for civil liberties, and respondents used a Likert scale. He used a factor analysis and cut the list of 12 down to 6, and the resulting summative scale had an alpha of only .62. Typical items were “Judges should punish criminals more severely” and “Any man who insults a policeman has no complaint if he gets roughed up in return” (p. 540). Gerbner et al. (1980) used three items from a national sample conducted by the National Council on Aging. These items asserted that the number, the health, and the longevity of older people are declining. They factor-analyzed the three and concluded that they belonged on the same dimension, but the alpha was only .56 which they concluded to be “moderate but acceptable” (p. 46). In 1980 Gerbner et al. provided evidence of reliability (combining three items to form an index of interpersonal mistrust), but the alpha was only .68. Signorielli (1990) reported low alphas on her mean world index (.67) and anomie index (.56), but concluded that these scales were reliable.

A third problem with scaling involves questionable practices. Typical of early cultivation research is the use of a cultivation index computed from four items such as (1) During any given week, what are your chances of being involved in some kind of violence? About 1 in 10? About 1 in 100?, (2) What percent of all males who have jobs work in law enforcement and crime detection? One percent? Five percent?, (3) What percent of all crimes are violent crimes like murders, rape, robbery, and aggravated assault? Fifteen percent? Twenty-five percent? and (4) Does most fatal violence occur between strangers or between relatives or acquaintances? (Gerbner et al., 1977). Besides the fact that both answer choices are wrong, the authors do not tell us how they constructed the index. We are left to assume that they may have added one point for every television world answer so that a score of 4 would indicate a perfect selection of television world answers and therefore provide maximum evidence of cultivation. If they in fact did this, they need to provide evidence that the scale is internally consistent, that is, that each of the four items does in fact contribute to such a scale, but we are given no reliability data. Later in the same article they mention constructing a mean world index from three items, again without explaining how they did so or how reliable their index is.

Hawkins and Pingree (1981a) used the same four mean world items and summed them to make a single index despite saying that they found the answers to the four items to be unrelated to one another. They dismissed this by saying that the “questions cover discrete bits of information about the
prevalence and nature of violence in society. As such, there is no particular reason why a person giving the TV-biased response to any one of these should necessarily also give the TV-biased response to any other” (p. 293). This statement is indeed troublesome. If they are truly unrelated conceptually, then they are not alternative items measuring the same concept, nor are they measuring related dimensions of the same construct.

Nature of the Relationship

There is ample evidence in the literature for a cultivation effect, that is, a relationship between how much television people watch and their likelihood of choosing the television world answer. This was the conclusion of a major review of the literature by Hawkins and Pingree (1982), who observed “most studies show evidence for a link, regardless of the kind of social reality studied,” and the topics include “prevalence of violence, family structures, interpersonal mistrust, fear of victimization, traditional sex roles, family values, images of older people, attitudes about doctors, and concern about racial problems” (p. 237).

Since the time of that review, at least 40 cultivation research studies have been published. There continues to be support for a cultivation relationship on feelings of anxiety, fear of victimization, and belief in a mean world (Morgan, 1983; Ogles & Sparks, 1989; Perse, Ferguson, & McLeod, 1994; Signorielli, 1990; Weaver & Waskshlag, 1986; Zillmann & Waskshlag, 1985). Also, there is support for a cultivation relationship on attitudes about approval of police brutality and bias against civil liberties (Carlson, 1983); political attitudes (Gerbner et al., 1982; Morgan, 1986); feelings of higher risks (Gunter & Wober, 1983); alienation and gloom, and anomie (Signorielli, 1990); loneliness, boredom, and depression (Morgan, 1984); estimates of doctors, lawyers, and illegitimate children of both males and females (Carveth & Alexander, 1985); black group identification and black separatist perspective (Allen & Hatchett, 1986); sex role stereotyping and sexuality stereotyping (Preston, 1990); and estimates of crime, working women, affluence, and health (Potter, 1991a).

Critique of Research on the Cultivation Relationship

Critics have been bothered that evidence for cultivation was too weak to conclude that such an effect existed (Doob & Macdonald, 1979; Hirsch, 1980, 1981a, 1981b; Hughes, 1980; Wober, 1978). They said that even by social science standards the magnitude of the coefficients is too low to conclude that television viewing is a useful predictor of cultivation estimates and beliefs. When Pearson correlation procedures are used, the resulting coefficients are typically below .15 and very rarely exceed .30. This means that the exposure variable usually predicts less than 3% of the variance in the cultivation indicator.

The theorists defended themselves from these criticisms by saying that the effect, while weak, is persistent (for example, see Gerbner et al., 1981). But critics maintained that the evidence for the cultivation relationship might be spurious for two methodological reasons: (1) The effect is dependent on a variety of “third” variables, and that when the influence of these other variables is controlled, the evidence for cultivation virtually disappears, and (2) the relationship might not be linear, so the use of statistics such as gammas or Pearson correlations is faulty. Each of these issues is examined below in light of the research of the past decade.

Control variables: When control variables are used in cultivation analysis, the resulting coefficients are almost always smaller than the coefficients computed without controls. Most researchers who use controls report that the reductions in coefficients are typically very small and do not change the interpretation of their results. For example, Gerbner et al. (1977) report four correlation coefficients ranging from .07 to .21 with a median of .16, and the magnitude of these was reduced to .06 to .18 with six simultaneous controls and a median of about .10 (Gerbner et al., 1977). Also, Hawkins and Pingree (1980) reported that the correlations remained relatively unchanged with controls for grade, sex, perceived reality media studies, newspaper reading, current events knowledge, and SES, either applied individually or simultaneously. Morgan (1984) found that heavier viewers felt more lonely, bored, and depressed, and these relationships held up in 11th order partials for education, income, time on hobbies, and worry index.

Carlson (1983) found cultivation evidence among sixth through twelfth graders on support for civil liberties, and these coefficients held up even with controls for grade, occupation, socio-orientation (degree of family authoritarianism), sex, hours of viewing total TV per week, reading habits, and parents’ favorite programs. Gross and Jeffries-Fox (1978) report a relationship between TV viewing and sexist views among children, and their correlation coefficient of .14 held up even with controls for grade, IQ, father’s occupation, and father’s education. Gerbner et al. (1980) report a cultivation relationship regarding conceptions of the elderly which held even under controls for education, income, sex, and age. And Preston (1990) found that there was a cultivation relationship among males between amount of exposure to pornography and sex role stereotyping and sexuality stereotyping. The coefficients remained almost unchanged even when controls were exercised individually or in total for age, income, strength of religious convictions, political/social orientation, GPA, daily TV viewing, current involvement of sexual relations, and satisfaction with current sexual relations.

Sometimes, however, the use of control variables results in a confusing pattern of results. For example, Gerbner et al. (1978) reported 13 gammas of .03 to .49 with a median of .14; when controls were exercised there was no...
consistent pattern with some gammas increasing and others decreasing even to the point of zero or becoming negative relationships, and there were no consistent patterns by control variable, that is, there was no one control (among age, sex, and education) that exhibited a consistent effect. In a later study, Gerbner et al. (1979) report eight gammas ranging from .17 to .52 with a median of .25; again no consistent patterns showed up with control variables.

Lack of Linearity: Is the cultivation relationship linear? Hirsch (1980) raised this question more than a decade ago when he reanalyzed the same data looked at by Gerbner and associates. Hirsch reported that the cultivation pattern across viewing groups appeared to be curvilinear. He found that non-viewers of television exhibited the highest degree of fearfulness, alienation, and anomie— all characteristics that Gerbner and colleagues claim are cultivated by television.

Despite the possibility that the cultivation relationship might be nonlinear, most researchers have ignored this finding and have not bothered to test for linearity and instead continued to compute Pearson correlations. Some have dropped back to using non-parametric statistics, but this lower level form of analysis also produces misleading results, because it ignores the nature of the relationship while purporting to describe it.

In a recent test of linearity, Potter (1991b) found that virtually any set of results could emerge from the same data merely by moving around the cut points when constructing groups. Starting with a continuous distribution, he used four different methods of transforming the distribution into categories (three equal sized groups, three groups by standard deviation cut points, five equal sized groups, and nine equal sized groups). The means of cultivated perceptions were not even monotonic across most of these groups, and the cultivation relationship (correlation between amount of viewing and degree of cultivated perception) was never linear on any of his measures (fear of victimization, affluence, stability of relationships, and health). It is possible, of course, that the cultivation relationship in this data set was anomalous (although the researchers found the standard weak positive bivariate relationships while using Pearson r’s), but we cannot know that until other researchers provide similar tests on their distributions.

It is clear that the practice of categorizing continuous distributions in cultivation research is not a neutral element in the analysis; instead it can react with the measures and often does, thus resulting in spurious findings. At best, it glosses over the complex pattern between the variables to focus on summary coefficients that indicate a weak positive relationship. These coefficients mislead us into believing that the relationship is a linear, weak, positive one when clearly it is not.

The results are rendered even more misleading when researchers use categorical measurement or analyses techniques without providing a compelling reason for the selection of their cut points. When they arbitrarily partition the curved relationship into discrete segments, some of those segments show a positive relationship, some a negative, and some no relationship. If the cut points are altered, then relationships within the categories can easily change. So which set of categories is best? Researchers must demonstrate that their cut points have conceptual significance. Why should a person who views two hours of television per day be regarded as a light viewer when a person who watches three hours is not? It is not sufficient to answer this question with something like “We are interested in comparing beliefs across relative levels of viewing.” If we are to accept this answer as evidence that there is a satisfying difference between two and three hours of viewing, then why not also say that there is a difference between three and four hours, which are usually stuck together in a single category of medium viewing? Also, it is not a satisfying argument to say that they are interested in relative comparisons among people at different viewing levels, because the use of continuous distributions would provide a much more powerful comparative test. Why spend effort quibbling about where the cut points should be? Instead, why not simply show the relationship plotted? Describing the shape would provide far more insight into the relationship than would forcing a linear based test.

While a graphic plotting of the relationship would be a step in the right direction, it will not serve as a complete solution to the problem of determining why evidence for the cultivation effect is persistently weak. Even when the relationship is plotted, researchers will have to test various prediction equations to provide a parsimonious description of the relationship. Should the prediction line be exponential, bell shaped, S-curve, sine wave, or something else (see Figure 2)? Researchers will need to examine the plots and try to determine what shaped line would best explain the scatter of points. Then they will need to test their postulated line in a multiple regression procedure. Potter (1991b) ran such a test on simple lines (total viewing which is the linear solution; total viewing squared and the square root of total viewing which are exponential solutions; reciprocal of total viewing; and log of total viewing) and found that a line that was the reciprocal of amount of viewing was the best line in predicting three of the four cultivation estimates. None of the lines explained much variance, however. Perhaps more complicated lines (such as an S-curve or a sine wave) are needed. These will be more difficult to test, because they would require additional terms in the predictive equation in order to plot each bend in the line. For example, the line in Figure 2 (a) is \( Y = a + bX^2 \). The line in (b) is more complicated and requires a formula such as \( Y = a + bX - cX^2 \). The more bends in the line, the more involved the formula. If each of the two variables (television viewing
and cultivation measure) in a plot have many values, then there is room for many bends in a predictive line, and the number of alternative lines available as a possible solution becomes very large. Thus testing of lines can become very time consuming and complicated.

There is also another factor that adds to the complex nature of testing for predictive lines, and that is the possibility of contingent relationships in the plot. In Figure 3, the top graph appears like a random pattern — no line could be plotted in a fashion to predict most of the variance. But what if the points above the dotted line represented the measures from adults and the points below the line represented children? A convex bow shaped curve could be plotted as a fairly good predictor for adults, showing a stronger cultivation effect at mid levels of TV viewing than either at low or high levels. In contrast, a concave bow shaped curve could be plotted as a fairly good predictor for children, showing a stronger cultivation effect at low and high levels of viewing.

Figure 2. Illustrations of non-linear relationships in cultivation analysis.

Figure 3. Illustrations of non-linear contingent relationships in cultivation analysis.
The plotting of scatter points and looking for (and testing) predictive lines is not an easy solution to the problem of testing for cultivation (for a discussion of this question, see Bauer & Fink, 1983). But this methodological strategy would at least keep researchers focused on the nature of the relationship. While this strategy would lead to a greater potential for extending our knowledge about the cultivation effect, at the tactical level, researchers will have a much more difficult task in conducting meaningful analyses of the data.

**Sampling**

Relative to other methodological problems, sampling is a strong point in cultivation research. Of course there are examples of non-probability type samples, such as convenience samples of college students (Bryant, Carveth, & Brown, 1981; Carveth & Alexander, 1985; Ogles & Hoffner, 1987; Ogles & Sparks, 1989; Perse, 1986; Potter, 1986; Weaver & Wakhlag, 1986), convenience samples of students in elementary and high schools (Alexander, 1985; Carlson, 1983; Gerbner et al., 1979; Gross & Jeffries-Fox, 1978; Hawkins & Pingree, 1980, 1981; Hawkins, Pingree, & Adler, 1987; Jeffries-Fox & Signorielli, 1979), and other convenience type samples (Gunter & Furnham, 1984; Rubin, Perse, & Taylor, 1988).

But there are many more examples of careful, representative samples. There is a good deal of research that generated data bases from telephone surveys using random digit dialing in Midwestern cities (Hawkins, Pingree, & Adler, 1987; Ogles & Sparks, 1989; Perse, Ferguson, & McLeod, 1994; Rouner, 1984) and in mid-Atlantic states (Einsiedel, Salomone, & Schneider, 1984; Perse, Ferguson, & McLeod, 1994; Stroman & Seltzer, 1985). There was a mail survey randomly drawn from voter lists (Volgy & Schwarz, 1980) and a door-to-door interviewing procedure was used with a random sample in Toronto, Canada (Doob & Macdonald, 1979). There are also many examples of secondary analysis of data from nationwide probability samples: NORC General Social Surveys (Gerbner et al., 1977, 1978, 1980, 1982; Morgan, 1983, 1986; Signorielli, 1990); the American National Election Study (Gerbner et al., 1978); Britain’s Independent Broadcasting Authority data based originally gathered by Gallup (Wober, 1978); National Council of Aging’s “Myth and Reality of Aging” survey (Gerbner et al., 1980); Roper data (Morgan, 1984; O’Keefe, 1984); Program for Research on Black Americans (Allen & Hatchett, 1986); and a regional sample from the Cincinnati Area Project (Fox & Philliber, 1978). Potter (1986, 1990, 1991a, 1991b) gathered data from a laboratory school composed of elementary and high school students selected from the surrounding county so as to construct a representative sample.

This proportion of probability to non-probability samples compares favorably to the balance found in social science research in general over a recent 25-year period, where less than 40% of that empirical research was found to use a probability sample or population study (Potter, Cooper, & Dupagne, 1993).


Many published articles reported the use of more than one sample (Gerbner et al., 1977, 1979, 1980; Hawkins, Pingree, & Adler, 1987; Morgan, 1986; Perse, Ferguson, & McLeod, 1994; Potter, 1986; Signorielli, 1990). Also, several studies conducted three-year panel studies to allow for some longitudinal analyses (Jeffries-Fox & Signorielli, 1979; Potter, 1991c). These features of multiple samples allow researchers to cross check the validity of their measures and findings as well as the reliability of their measures.

**Recommendations**

When the empirical findings of cultivation studies to date are simply summarized, the conclusion is that there is a persistent pattern of a positive relationship between amount of exposure to television and cultivated beliefs or estimates. This is the same conclusion reported over a decade ago by Hawkins and Pingree (1982).

A major reason for the same conclusion is that there has been very little change in methodological practices. And therein lies the problem. The analysis above has identified a range of methodological flaws in cultivation research. Of course, it cannot be stated with certainty that had these flaws not existed, the conclusions of the research would be substantially different. Nevertheless, there is reason to be hopeful that avoiding these flaws in the future can lead to some fresh insights about the cultivation effect.

Below are six methodological recommendations for cultivation research. Each one has been synthesized from the patterns of problems identified in the literature. This set serves as a summary for the critical evaluation of the literature that has gone before.

1. **Recognize the Relationship as Non-Linear**

The most critical methodological problem with cultivation analysis is the implied acceptance of an assumption that the relationship is linear. This absolves researchers from justifying the cut points they use when transforming continuous distributions on their variables into categorical ones. It also gives permission for researchers to use categorical measures. Both of these practices
parameters. There may be a good reason for doing this, but nowhere in the
theory is there a clearly articulated rationale. And nowhere in any of the
research is there a justification for why respondents should not be allowed to
volunteer their own estimates without being directed to choose between two
false answers.

3. Use of Test Variables

Test variables help determine whether a bivariate relationship is
spurious because of the influence of some “third” variable. Testing for these
variables is especially important with the cultivation relationship, which
over the long term is sensitive to the influence of many “third variables.”
While the use of these “third variables” generally strengthens the research,
sometimes their use is not conducted properly, and the resulting insights are
questionable.

First, there is a problem in deciding what to include as a control
variable. The research has taken on an exploratory feeling on this point,
because the theoreticians provide very little guidance about which variables
should be tested with certain cultivation indicators. Gerbner often
acknowledges the importance of control variables, but he does not provide a
good conceptualization of which variables to use in which situations, and it
is difficult to tell what variables are important from examining his analyses
of data.

Second, the theory provides no guidance as to when a “third variable”
should be used as a control and when it should be used as a contingent
variable. Control variables are those whose influence on the bivariate
relationship are mathematically removed through partial correlation.
Contingent variables are used to construct sub-groups of respondents to test
for differences in the relationship across groups. This decision of whether to
treat a variable as a control or as a contingent variable can be made on two
grounds: conceptual and mathematical. Continuous variables can be treated
as either control or contingent variables depending on how they are
conceptualized. If they are conceptualized as “nuisance” variables whose
influence should be set aside, then they should be treated as control variables
in the methods. But if their influence on the relationship is an important focus
(for example, a mainstreaming or resonance effect), then they should be
treated as a contingent variable.

From a mathematical point of view, categorical variables should be
treated as contingent variables, because their mathematical properties do not
lend them to interpretation in partial correlations or multiple regressions. At
first, this point might not appear obvious. Categorical variables (such as
gender and race) are frequently used as predictors in multiple regressions.
Look at this more closely through an example: If we gathered data on
respondents’ race such that white were coded as a 1, black as a 2, Hispanic as
a 3, and other as a 4, the level of measurement would be nominal, and it would

2. Justifying the Television World Answer

Cultivation theory requires an a priori identification of the television
world in contrast to the real
world. In practice, researchers
have either ignored the
requirement to justify the
selection or have provided
faulty reasoning as a
justification. This is a serious
shortcoming. This criticism was
first raised by Newcomb (1978),
and it has not been addressed
adequately by the theoreticians.
Furthermore, very seldom do
authors of empirical work
provide any kind of a rationale for selecting a television world answer.

Such a rationale would start with the results of a carefully conducted
content analysis; however, many studies do not even provide this. But even
with content analysis findings, there is a need to examine the qualitative
context of acts in order to justify the television world parameters. For
example, the frequency of a lot of crime on television does not by itself
provide a justification for saying that the television world is a criminal world.

Also, there has been a very curious methodological practice of providing
closed ended forced choice answers between two incorrect real world
parameters.
be a serious mistake simply to enter race (with its four values) into a regression equation as a predictor variable. Instead, we need to create four dummy variables — one for each of the values. Therefore, the effect of race on the predictive equation is not examined directly; it is examined indirectly by providing a predictive equation for each of the four dummy variables. This, in essence, is a contingent analysis, because it allows us to focus attention on the differences across the four equations. If there are no differences in those equations, then the predictive equation (composed of the other predictor variables) is not sensitive to race. If the equation is different, then there is evidence of a contingent relationship, that is, the relationship between the predictors and the criterion variable (usually the cultivation measure) varies across groups.

Providing a strong rationale for test variables is especially important when dealing with contingent analyses. The only major conceptual development in the theory since its beginning has been the ideas of mainstreaming and resonance, both of which require a contingent analysis to test. But in order for the test to be meaningful, there must be a strong rationale for selection of the contingent variables and there must be a clear articulation about how the relationship should vary across contingent groups, but this is rarely accomplished. For example, Gerbner et al. (1980) reported that they found evidence of a mainstreaming effect when they observed a difference on the mean world index between whites and non-whites. The white group displayed a cultivation coefficient of .23, while the coefficient of non-whites was -.10. They explained this difference (in a post hoc discussion) by saying that whites are less likely to have a mean world attitude so the TV world has a stronger effect on them in bringing them into the “mainstream” of the television world which is violent. In contrast, non-whites live in a more violent real world so they have a more extreme view than that presented on television, so heavy TV viewing would reduce their feeling of a mean world. It is very difficult to accept their logic. Perhaps the white part works, but the non-white part is absurd. The only way we could support the mainstreaming explanation of the non-white viewer is to say that non-whites hold a mean world attitude so strong that the violent world of television (where two-thirds of all characters are involved in violence) is peaceful by contrast, so that heavy viewing serves to remove the viewer from the violent real world and bring him/her into the mainstream of the comparatively peaceful TV world! The evidence for mainstreaming is mixed even within that one article. For example, they found a relatively strong (.14) relationship with anomie among higher educated (some college) viewers. But they did not find a negative relationship among high school dropouts.

The problem with using mainstreaming is that the researcher must do more than simply point to a contingent pattern and evoke mainstreaming as an explanation. The mainstreaming explanation only works when researchers provide a strong rationale (preferably in an a priori fashion) to justify which groups are outside of the mainstream. Then the researcher must look at the overall pattern across groups and not limit the explanation to a single anomalous group.

Third, there are problems with the computation of the influence of control variables. For example, in one typical study, Gerbner et al. (1977) used a set of control variables with widely differing metrics. They compute Pearson product moment correlation coefficients to express the relationship between a mean world scale (range 0 to 3) and exposure (range 1 to 3). With such a truncated range of values on both variables there is very little variance, and the resulting coefficients should be corrected for attenuation. Then they use partial correlation coefficients to control for sex, age, education, income, newspaper reading, and church attendance. Sex, of course, has a binomial distribution; we do not know about the other variables, although given the researchers’ practice of turning continuous distributions into categorical ones, these other variables might also have binomial distributions. If they do, then the partial correlation test is not a useful one. If instead there is a mix of distributions, then the results should be highly suspect, especially given the fact that we are not beginning with very strong coefficients. In other words, when we begin at a position of not being able to predict much variance in the relationship between two variables, and we test to see if the variance that we have predicted is spurious because of the influence of control variables, then the use of control variables with very little variance sets up a condition where it is almost impossible to find a significant change in the partialled coefficients compared to the zero order coefficients.

Given the problems outlined above, it is not surprising that control variables have not been found to be good test variables. Without a clear conceptualization about what should be a test variable and what effects they should have on the bivariate cultivation relationship, it is possible that some of the test variables may have served to increase the magnitude of the relationship while others have served to reduce it, thus making the net effect one that obscures the relationships rather than clarifying our understanding of the complex nature of the interactions.

4. Stronger Research Designs

Several researchers have conducted experiments to test the cultivation effect (Bryant et al., 1981; Ogles & Hoffner, 1987). With many research topics, experiments are superior to surveys because researchers have more control over their subjects and can design experiences that allow them to make causal inferences in their results. But the situations and controls must conform to the
cultivation effect perspective in order to be a useful test of this theory, and this could easily be a problem. For example, Ogles and Hoffman designed a classic experiment, but it is highly questionable whether the results reflect a test of cultivation, because the researchers exposed their subjects to only five films over a two-week period. Although this treatment is more extensive than many media effects experiments, it hardly simulates a cultivation-type influence. If the authors wanted to examine how certain types of programs have a short-term shaping influence on the overall cultivation process, they should have taken measures of cultivation indicators before and after their treatment, but they did not. Bryant et al. (1981) avoided this type of problem by assessing the trait anxiety of their subjects before running an experiment. This allowed them to track the level of anxiety, and they found that the low anxiety group’s scores increased during the viewing while the high anxiety group’s scores decreased.

The experiment has a place in cultivation research, but its use must be recognized as a test of short-term influences within the overall long-term process of cultivation. Of course, it is possible that any observed short-term influences may stay with the respondents and thereby have a long-term effect. Therefore, it is important that experimental researchers continue to measure their respondents years after their treatments to determine if the effects they generate in the laboratory remain the same, decay, or become magnified over time.

A stronger design would be an ethnographic approach where a researcher would observe people watching television over a long period of time. The researcher would document how people interact with the television receiver and other human beings as well as how those people would exhibit long-term cultivation effects. This, of course, would require an enormous commitment of time. A more efficient, but still time-consuming, method would be in-depth interviewing of individuals in time series. Researchers would follow a panel of individuals by talking with them at six-week intervals over a period of several years.

The survey method has been useful in the past, and it will continue to be useful if applied properly and if its use were extended to more sophisticated designs such as those using cross-lagged analysis, which has been employed successfully in research on the long-term effects of television viewing on behaviors (Milavsky, Kessler, Stipp, & Rubens, 1982). The use of a wider range of methods would provide the opportunity to triangulate findings and determine to what extent our knowledge of cultivation is sensitive to particular forms of data gathering.

5. Quality of Measures

The measures used in cultivation research are subject to the same criticism as all measures used in social science research. Cultivation researchers need to be more careful in meeting the basic psychometric criteria of validity, reliability, and usefulness when measuring the two primary constructs: TV exposure and cultivation indicators. When scaling procedures are used, researchers should be more careful about making the case that their measures and scales achieve appropriate psychometric quality. This means that researchers need to present clear evidence that their scales are reliable, valid, and useful. The latter two of these are almost totally ignored as researchers seem to feel that the case for these is obvious. There is greater attention paid to reliability, but often the case is made only in a partial manner.

There are also some special psychometric problems in cultivation research. For example, scholars have raised some serious questions about the validity of self-reported exposure measures. Much has been written about the relative merits of diaries, telephone coincidental methods, and people meters, but the conclusion is that the results of the various methods do not always correlate very highly (for example, see Beville, 1985). Also, viewers might not be able to provide valid cultivation measures. Rubin, Perse, and Taylor (1988) report evidence that respondents who estimate high on their TV exposure might also be overestimators on the cultivation indicators, and the resulting relationship would then be a reflection of the characteristic of overestimation, not a cultivation effect. They caution researchers to minimize “response bias by using positively and negatively phrased items” (p. 125).

6. Sampling

In several areas, cultivation is still in an exploratory phase. In the exploratory areas of psychometrics (testing different measures of exposure and different types of cultivation indicators) and model building (determining the psychological processes that underlie cultivation), it is not crucial that the samples be probability ones. Of course, it would be better if probability samples were used, but the value of exploratory findings can be quite high with non-probability samples.

Once the research establishes firmer guidelines for measurement and analysis, then the focus should shift to the task of determining the extent of the cultivation effect on different types of people. For this work, probability samples are, of course, required. But this may be very difficult to achieve as successfully as in the past. Many of the cultivation research studies that have used probability samples in the past were secondary analyses of existing data bases that were gathered for other purposes than to test the cultivation theory. If researchers are to continue using such data bases, they will need to influence those large data gathering organizations to alter their items so that the resulting data will conform to the developing standards of this theory.

Conclusion

In summary, cultivation researchers need to provide a greater degree of care in designing their measures and analyses. When they do so they will realize that it is better to use higher level of measurement; preserve continuous
distributions in their scaling; use more sophisticated statistical techniques (such as multiple regressions with curvilinear predictions, rather than gammas or Pearson's r's); provide stronger rationales for "third variables" that illuminate their individual and simultaneous influences; and, above all, keep focused on providing more complete descriptions of and explanations for the shape of the relationship. The cultivation effect is a complex non-linear relationship that is influenced to differential degrees by different "third variables" at different points in the curve. Until cultivation researchers begin building this complexity into their designs, the research will stay mired in its tradition of faulty methodological practices that can only contribute another list of potentially spurious coefficients.

References


