OFFICE OF SCALE RESEARCH

Technical Report
#9201

A Review of the Reliability & Validity of Multi-Item Power Scales

by

John T. Drea
Dr. Gordon C. Bruner II
Paul J. Hensel

1992

About the Authors

John T. Drea (DBA candidate, Southern Illinois University; MBA, University of Notre Dame) is the Assistant Dean of Instruction at John Wood Community College in Quincy, Illinois. His research interests are in the field of sales promotion and consumer behavior. His work has previously appeared in the Proceedings of the Southern Marketing Association.

Gordon C. Bruner II (Ph.D., University of North Texas) is Associate Professor of Marketing at Southern Illinois University and Director of the Office of Scale Research. Among his articles are those appearing in the Journal of Marketing, the Journal of the Academy of Marketing Science, the Journal of Retailing, and the Journal of Marketing Education.

Paul J. Hensel (Ph.D., University of Houston) is Associate Professor at the University of New Orleans and has published articles in the Journal of the Academy of Marketing Science, the Journal of Advertising, the Journal of Business Ethics, the Journal of Health Care Marketing, and several other journals and conference proceedings.

A REVIEW OF THE RELIABILITY AND VALIDITY OF MULTI-ITEM POWER SCALES

Abstract

Twenty nine multi-item scales used to measure power, power sources, and/or exercised power in a sales or channel setting are reviewed. The procedures for assessing the reliability and validity (convergent, discriminant, and nomological) of each scale are discussed and reviewed. Findings indicate strong reliability for all of the scales. Problems in the establishment of validity for some of the coercive and legitimate power measures are noted, along with recommendations for selection and future research of power scales.

A REVIEW OF THE RELIABILITY AND VALIDITY

OF MULTI-ITEM POWER SCALES

The measurement of power is central to understanding the behavior of organizations and individuals. *Power* is commonly defined as "the ability of one individual or group to control or influence the behavior of another" (Hunt and Nevin 1974). Power has served as an important construct because of its hypothesized relationship to other variables such as satisfaction, role performance, and conflict. However, despite its importance, measures of power have received relatively little critical comparison and analysis during recent years. Additional research is needed into creating and testing valid and reliable multi-item scales for measuring power (Podsakoff and Schriesheim 1985).

The purpose of this research is to examine the measurement of power constructs using multi-item, multi-point Likert-type scales, with a focus on the procedures used to assess scale validity and reliability. The convergent, discriminant and nomological validity of each of the power constructs will be examined using the criteria discussed by Peter (1979, 1981), Churchill (1979, and Peter 1984), Gerbing and Anderson (1988), and Bagozzi and Yi (1991). Conclusions will be drawn regarding the reliability and validity of each scale, and recommendations will be made regarding preferred measures and further scale development.

The origins of most research into the power construct can be traced to the work of French and Raven (1959), who identified five bases (sources) of social power:

Reward power source:	The perception of an individual or organization (B) that another individual or organization (A) has the ability to mediate rewards for B.
Coercive power source:	The perception of B that A has the ability to mediate punishment(s) for B.
Legitimate power source:	The perception of B that A has a legitimate right to prescribe behavior for B.
Referent power source:	The extent to which B identifies with A.
Expert power source:	The perception of B that A possess special knowledge or expertness.

These five sources of social power serve as the foundation for most of the model development regarding the construct of power in marketing.

Much of the development of the original hypotheses regarding the effect of power sources on actions in a marketing channel is credited to El-Ansary and Stern (1972), Hunt and Nevin (1974) and Lusch (1977). Subsequent research has also resulted in revised depictions of power by Etgar (1978), John (1984), and Gaski (1986).

Other variations from the original five bases of social power are related to the use of perceived versus exercised measures. *Perceived* measures have been popular in power research (e.g., El-Ansary and Stern 1972, Hunt and Nevin 1974, Michie and Sibley 1985). In particular, Raven (1965) explicitly defines power as a perceived measure and not as an objective phenomenon.

Much of the marketing channel literature has commonly accepted that the perception of power may have more influence on channel member behavior than the actual exercise of power. Gaski and Nevin (1985) noted that the *exercise* of a coercive power source, however, has a stronger effect on satisfaction and channel conflict than the mere presence of a coercive power source. In comparison, the exercise of a reward power source was found to have only a minor effect on satisfaction and channel conflict.

METHODOLOGY

A review was conducted of all of the articles published in Journal of Marketing Research, Journal of Marketing, and Journal of the Academy of Marketing Science, and Industrial Marketing Management for a ten year period from 1980 to 1989. These particular journals were selected for the review due to being the most well known outlets in the field of marketing for scholarly articles which investigated the power construct during that decade. Additionally, selected influential articles on the power construct which appeared in other sources (Journal of Applied Psychology and the Journal of Personal Selling and Sales Management) were also reviewed. A review of more than 600 articles revealed that nine journal articles measured either power or power sources using multi-item scales, as suggested by Churchill (1979). Some of the studies measured the *perception* of power sources using the five original sources of power (John 1984; Comer 1984; Michie and Sibley 1985; Frazier and Summers 1986; Kohli 1989; Hinkin and Schriesheim 1989), while others

measured the *exercise* of power (Gaski 1986; Butaney and Wortzel 1988; Gaski and Nevin (1985) measured both exercised and perceived measures of power in the same instrument.

Each of the power scales reviewed used a multi-item scale to measure either one of the five sources of power delineated by French and Raven (1959), or a broad measure of power, such as "manufacturer power" (Frazier and Summers 1986). Measures excluded from the present analysis include all single item measures (see Podsakoff and Schriesheim 1985 for a review of these), and measures which separate and/or combine the sources of power differently than in the original reward/coercive/ legitimate/referent/expert sources (e.g., Schul, Remington, and Berl 1990, who contingent and non-contingent reward and punishment behavior).

To critique the power scales selected, several criteria related to the reliability and validity of the multi-item measures were collected and analyzed. In particular, the number of scale items, scale points and sample size are reviewed because these items were cited by Churchill and Peter (1984) as having a significant relationship with scale reliability. The procedures used by each author to establish the convergent, discriminant, and nomological validity of each measure are also included in the review. Validation procedures used in each article are compared with those discussed by Bagozzi and Yi (1991). Validity criteria and procedures discussed by Heeler and Ray (1972), Churchill (1979), and Peter (1981) are also used to assess each of the reported validation procedures.

RESULTS

Reliability

In the present research, reliability is defined as "the degree to which measures are free from error and therefore yield consistent research" (Peter 1979: p. 6).

Procedures for Assessing Reliability. A summary of the reliability findings is provided in Table 1, which also highlights the number of scale items used in each measure, the sample size, the reliability measure used (Cronbach's alpha), and findings.

****

Insert Table 1 about here

****

Of the 29 power scales reviewed, 28 reported evidence of reliability through the use of Cronbach's alpha. This procedure provides a relatively simple and well accepted assessment of the internal consistency of a given measure.

Potential Influences on Scale Reliability. Of the 28 power scales reporting reliability coefficients, sample sizes ranged from 147 to 251. There were three exceptions: Frazier and Summers (1986) used a sample of 435; sample B used by Hinkin and Schriesheim (1985) had 375 subjects; and Butaney and Wortzel (1988) used a sample size of 83. There was also a generally narrow range of the number of items (4-10) reported for each scale, with the only exceptions being the 15 item scales used by Gaski (1986, Gaski and Nevin 1985) to measure perceived and

exercised reward power, and a 17 item scale used by Butaney and Wortzel (1988) to measure distributor power. Finally, there was a relatively narrow range of the number of scale points, which ranged from 4-7 points with only two exceptions (Michie and Sibley 1985: 8 points; Frazier and Summers: 11 points). The relatively narrow range of sample sizes, scale items, and scale points used with each scale indicates that these factors alone are unlikely reasons for any given scale to perform more reliably in comparison to any other scale.

A correlational analysis of the 28 scales reporting reliability coefficients was conducted to examine any potential relationship between sample size, number of scale items and number of scale points. The results (see Table 2) failed to provide evidence for a significant relationship.

****

Insert Table 2 about here

****

As indicated by Churchill and Peter (1984), these items can potentially increase the reliability of a given scale. A suggested minimum ratio of 5-10 subjects per item for samples under 300 has been suggested (DeVellis 1991). While the overall correlations were found to be insignificant, one particular scale which may have been effected by these items is the distributor power scale (Butaney and Wortzel 1988), which possessed a sample size (83) insufficient for the large number of scale items (17).

Reliability Results. Satisfactory reliabilities (.70 and above, Nunnally 1978) were found for most power measures. In

particular, the reliabilities for the reward power scales (alphas ranged from 0.77 to 0.92, mean = 0.86), the referent power scales (0.81 to 0.88, mean = 0.85) and the expert power scales (0.77 to 0.90, mean = 0.83) were consistently high and in a relatively narrow range. The reliabilities for the coercive power measures reporting a reliability coefficient were lower and possessed greater variation (0.62 to 0.90, mean 0.76). The reliabilities for the legitimate power scales were more consistent but were also generally low (0.65 to 0.86, mean = 0.72). These findings suggest that the internal consistency of the measures of each of the above reviewed constructs is relatively consistent, regardless of the particular scale employed.

Convergent and Discriminant Validity

Churchill has defined convergent validity for a given measure as "the extent to which it (the measure) correlates highly with other methods designed to measure the same construct," while defining discriminant validity as "the extent to which the measure is indeed novel and not simply a reflection of some other variable" (Churchill 1979: p. 70).

Procedures for Assessing Convergent and Discriminant Validity. The procedures used to investigate convergent and discriminant validity are listed in Table 3. The power scales are compared on each of the three procedures below.

****

Insert Table 3 about here

****

As seen in Table 3, exploratory factor analysis is a common procedure for assessing convergent and discriminant validity of multi-item power scales. Exploratory factor analysis has proved useful as a preliminary technique for reducing the number of scale items and for confirming a researcher's hypotheses as to how items should group together (DeVellis 1991). However, exploratory factor analysis does not provide an explicit test of the unidimensionality of the construct (Gerbing and Anderson 1988), which is provided through confirmatory factor analysis procedures such as those available through the LISREL program (Joreskog and Sorbom 1984). While several of the scales (five scales by John 1984, two scales by Kohli 1989) employed exploratory factor analysis, only Hinkin and Schriesheim (1985) and Gaski (1986) appear to have used a confirmatory procedure (LISREL) to test a structural equation model and to assess discriminant validity.

The scales used by Gaski (1986, Gaski and Nevin 1985) made use of portions of the multitrait-multimethod (MTMM) approach to assess convergent and discriminant validity as outlined by Campbell and Fiske (1959). The validation procedure used by Gaski (1986, Gaski and Nevin 1985) compared the reliability coefficient for each power scale measure with the correlations between the measure of interest and other power scale measures, and used alternative measures of the same construct. As an additional check of discriminant validity, Gaski (1986) also used LISREL in a procedure similar to one suggested by Bagozzi and Phillips (1982) to examine whether the five power source

constructs were perfectly correlated. In regards to convergent validity, Gaski and Nevin (1985) assessed convergent validity for their measures of reward and coercive power through correlations with alternative measures of the same construct.

Hinkin and Schriesheim (1989) used both exploratory and confirmatory factor analysis, plus correlational analysis to assess convergent and discriminant validity. Confirmatory factor analysis was used to assess the goodness of fit and unidimensionality of the five scales. Exploratory factor analysis was used to assess discriminant validity by determining whether items from the five power bases would cross load on non-power base items (satisfaction and commitment).

The Attributed Power Index (API) scales, originally developed by Holzbach (1974), were subjected to the full multitrait-multimethod validation procedure by Comer (1984). Correlations were calculated: a) between each of the multi-item measures, b) between each of the single item, alternative method measures, and c) between the single and multi-item measures. Reliabilities were calculated for each measure in the single and multi-item measures.

John (1984) and Butaney and Wortzel (1988) both used exploratory factor analysis to assess convergent and discriminant validity. Butaney and Wortzel (1988) also used correlational analysis to assess the relationship between the multi-item measure of manufacturer power and a single item measure.

Frazier and Summers (1983) used confirmatory factor analysis to assess discriminant validity by measuring manufacturer power

at the strategic and tactical levels of the organization. Convergent validity was assessed by correlating manufacturer power with a related measure intended to measure dependency.

Selection of Alternative Methods. A consistent issue for the convergent validation of many of the scales used by Comer (1984), Gaski and Nevin (1985), Gaski (1986), Frazier and Summers (1986) and Hinkin and Schriesheim (1989) surrounds the selection of appropriate alternative methods. One of the problems of using the MTMM matrix approach to validation is related to the development of different methods for assessing the same construct. Additionally, it is not enough that methods are superficially different: They must be maximally different methods that avoid the problems of shared method and method-trait variance (Peter 1981). While estimates of the method variance present can be calculated using procedures such as path analysis or confirmatory factor analysis, Peter (1981) has argued that "different forms of paper and pencil self-rating scales are clearly not maximally different methods" (p. 137). The use of such similar methods is more likely to produce an assessment of alternative form reliability than an assessment of convergent validity (Cohen 1979). Such an assessment is not undesirable, but it does not replace the need for providing a clear assessment of convergent validity. Thus, the use of similar methods or the use of multiple samples with the same scale items cannot be confidently used to substantiate claims of convergent validity.

Both Comer (1984) and Gaski (1986, Gaski and Nevin 1985) made use of alternative measures. The alternative scales used by

Comer (1984) are the single item power measures developed originally by Busch (1980). These single item scales were administered in the same manner as the multi-item API measure: a paper and pencil survey mailed to sales representatives. The alternative measurement instrument used by Gaski and Nevin (1985) consists of a series of Likert-type power measures, which are similar to the primary power measures. The alternative measures used by Hinkin and Schriesheim (1989) appear to be a repeated use of a single measure with multiple samples, rather than true alternative measures. Claims of convergent validity for these power scale measures should be viewed with varying degrees of caution.

Convergent and Discriminant Validity Results. Gaski (1986, Gaski and Nevin 1985) claim evidence of discriminant validity for each of their five power scales. The alpha coefficient for each measure was greater than the correlation of the measure with other measures, leading to the claim of discriminant validity. The confirmatory procedure used by Gaski (1986) also supports that the five power source constructs were not perfectly correlated with one another, and that discriminant validity for each of the five power source measures was achieved. Convergent validity is asserted by Gaski and Nevin (1985) for their measures of reward and coercive power on the basis of positive and significant correlations with alternative measures of the same construct. However, no alternative measures were reported for the legitimate, referent, and expert power source measures by Gaski (1986).

As indicated earlier, Comer (1984) used the full MTMM procedure to assess validity. His findings report a failure to establish convergent and discriminant validity for the coercive, reward, and legitimate power source scales, while such requirements were largely satisfied for the referent and expert power source scales.

The claim of discriminant validity for the scales used by Hinkin and Schriesheim (1989) is supported by factor loadings, indicating that conceptually distinct measures (satisfaction, commitment, and power) loaded on different factors, with no inappropriate loadings. The results provide limited support for convergent validity, with most items loading on a single factor (some of the expert power items in Sample B also loaded at .40 or greater on legitimate and referent power). However, using Churchill's definition, the use of a scale with multiple samples (as opposed to alternative measures of the same construct) appears to be more of a measure of reliability than convergent validity.

John (1984) provides limited evidence of convergent or discriminant validity. Factor analysis was used to determine a one factor model fit to the data for each scale. Factor loadings, however, are not reported.

Evidence of convergent and discriminant validity is provided for the Frazier and Summers (1986) measure of manufacturer power. Using confirmatory factor analysis, factor loadings of 0.73 or greater were achieved on each hypothesized factor, with no significant cross-loadings. The use of correlations between the

manufacturer power measure and a related dependency measure appears to be evidence of nomological validity, rather than an alternative measure of the same construct (as required by Churchill in his definition of convergent validity). On the whole, however, the findings generally support Frazier and Summers' claim of convergent and discriminant validity for the manufacturer power measure.

Butaney and Wortzel (1988) provide evidence of convergent and discriminant validity for their weighted measure of distributor power through factor loadings, and through a significant correlation between their measure of distributor power and a single item measure. The results support the claims of convergent and discriminant validity.

Kohli (1989) and Michie and Sibley (1985) both provide rather limited evidence of convergent and discriminant validity for their measures of referent and expert power (Kohli) and coercive power (Michie and Sibley). Exploratory factor analysis loadings were uniformly high (.59 or greater for each item in the four item scale, with no item loading higher than .32 on any other factor). However, the measures used by both authors lack an alternative measure for a more meaningful convergent or discriminant validity.

Nomological Validity

The assessment of nomological validity "entails investigating both the theoretical relationship between different

constructs and the empirical relationship between measures of those different constructs" (Peter 1981: p. 135).

Procedures for Assessing Nomological Validity. Nomological validity was investigated in 25 of the 29 power scales reviewed. Of these 25 investigations, all indicate some positive evidence of nomological validity by correlating the given power measure with another measure related by theory.

Table 4 highlights the procedures used to establish the nomological validity of the measure, indicating with what other measure(s) the scales of interest were correlated. While *satisfaction* has been commonly used as a construct for assessing the nomological validity of power scales, the operationalization of the satisfaction construct has taken many different forms. For example, Hunt and Nevin (1974: p. 189) define satisfaction as "what (the franchisee) would do `if he had to do it all over again.'" Comer (1984) measured seven dimensions of satisfaction for sales representatives.

****

Insert Table 4 about here

****

Since these measures and others assess different interpretations of satisfaction, a different means for assessing nomological validity was needed. For this reason, the correlations between the five French and Raven (1959) power source measures were collected and analyzed. Such correlations were reported by Comer (1984); John (1984); Gaski and Nevin (1985); Gaski (1986); and Hinkin and Schriesheim (1989). The

expected pattern of such correlations has been previously suggested by Raven and Kruglanski (1970), who reported that there should be positive relationships between the variables of expert, legitimate, and referent power sources. The basis for these propositions by Raven and Kruglanski (1970) is that the three power sources are considered to rely on the same mediating processes. Subsequent models of power by Hunt and Nevin (1974) and Gaski (1986), while differing in some important respects, have included these three power sources as dimensions of a larger, underlying trait. Raven and Kruglanski (1970) have also suggested that there should be a negative relationship between coercive power and expert, referent, and legitimate power. Table 5 indicates the reported correlations between each of the five sources of power.

****

Insert Table 5 about here

****

Nomological Validity Results. Among the present studies reviewed, each of the scales using satisfaction as a correlate reported that a nomological relationship was achieved.

Claims of nomological validity are supported for most of the measures, though not all. The referent, expert, and legitimate power scales used by Comer (1984), John (1984), and Gaski (1986) correlate with one another in a predictable pattern, with all correlations being positive and strongly significant. The consistency of these findings supports the claims of nomological validity for each of these scales.

Most of the reward power source scales were also found to be correlated in a predictable pattern. Each reward scale correlated positively and significantly with the other scales, with correlations ranging from r = .2 to .46. The exception was the reward scales used by Comer (1984), which were more strongly correlated with referent power (r=.73) and expert power (r=.57).

With regard to coercive power measures, Gaski (1986, Gaski and Nevin 1985) has noted that the exercise of a *coercive* power source produces a different effect on reward power than on other power sources. Gaski and Nevin (1985) have noted that the correlation between reward power source and coercive power source (both *perceived* measures) is expected to be positive, while a negative correlation is expected between reward power sources and *exercised* coercive power sources. Accordingly, the effects of the exercised coercive power source measure on reward power sources must be viewed as a different relationship than the effect of one power source on another.

The coercive power measure used by Comer (1984) is a part of Holzbach's (1974) API index and performs as expected when correlated with reward (r=.22) and legitimate (r= -.12) power sources. However, the API-based coercive power measure displays a very low and positive correlation (r=.02) with expert power source, when a negative correlation is expected. In addition, the coercive power measure used by Comer (1984) is correlated positively with referent power source (r=.25), while a negative correlation is expected. Accordingly, the nomological validity

claim of this coercive power scale cannot be confirmed on the basis of correlations with other power source scales.

The legitimate power source measure reported by John (1984) performs as expected in correlation with reward, referent, and expert power sources. However, the measure displayed a very low and positive correlation (r=.01) with a coercive power source measure when a negative correlation was expected. Accordingly, the nomological validity claim of John's (1984) legitimate power scale should be viewed cautiously.

As expected, the coercive power scale used by Hinkin and Schriesheim (1985) correlated positively in three samples with reward power and negatively with three separate measures of satisfaction, thereby supporting the nomological validity of the scale. However, the coercive power measure did not correlate as expected with legitimate power, yielding significant positive correlations in three samples (r = .20, .16, and .41) when a negative correlation was expected. Additional investigation is needed to confirm the nomological validity for the legitimate and coercive scales.

Extensive support is provided for the claim of nomological validity of the manufacturer power scale used by Frazier and Summers (1988). In an earlier test of this scale (Summers 1983), manufacturer power was shown to be highly correlated in an expected pattern with five related measures, including satisfaction (r=.28 and .57 for corporate and boundary personnel, respectively).

Both Butaney and Wortzel (1988) and Kohli (1989) do not appear to have specifically investigated the nomological validity of the distributor power or the referent and expert power scales, respectively. Correlations between power scales or between power scales and related constructs are not reported. Accordingly, there is little additional support for the nomological validity of these scales.

Nomological validity of the coercive power scale used by Michie and Sibley (1985) does not appear to have been specifically investigated by the authors. However, there does appear to be some support for its nomological validity. Coercive power was shown to be a significant and negative predictor of satisfaction (b= -.314, p=.005), as suggested by others (e.g., Burke and Wilcox 1971; Busch 1980; Hinkin and Schriesheim 1989).

It is important to note that a failure to confirm a claim nomological validity does not mean that nomological validity is absent. There are alternative explanations of why two measures may not correlate in an expected pattern, including the particular dimensions measured by both measures, as well as the reliability of the measures. It is for these and other reasons that it has been suggested that the nomological validity process include correlations with several measures that have been previously established as being construct valid (Churchill 1979), and not just one measure. The scales used by Comer (1984), John (1984), Gaski (1986, Gaski and Nevin 1985), Frazier and Summers (1986), and Hinkin and Schriesheim (1989) used correlations with several different measures to support a claim of nomological

validity. The use of multiple correlations with other established constructs increases the opportunity for the support of nomological validity for each of these scales.

DISCUSSION

In general, all of the scales were found to possess at least acceptable levels of reliability. All of the reviewed scales used to measure reward, referent, and expert power had high internal consistency (0.77 or greater), while all of the scales used to measure legitimate power sources were uniformly lower, though acceptable.

Most of the power scales addressed either directly or indirectly the issues of convergent and discriminant validity, with most using either one or more of the following: exploratory factor analysis, alternative measures of the same construct, or a version of the multitrait-multimethod (MTMM) matrix approach. Some cautions are noted, including the lack of a test of unidimensionality in exploratory factor analysis; the use of alternative measures for validation purposes when such measures are not maximally different; and the use of only part of the total MTMM matrix. The presence of such weaknesses does not refute the convergent and discriminant validity of the scales, but concurrently, does not provide strong evidence of the validity of the measures either.

Finally, nomological validity was discussed either directly or indirectly for twenty-two of the twenty-nine power source scales. Each of these twenty-two scales investigated and

reported some evidence of nomological validity by correlating the scale with another theoretically related measure, such as satisfaction, power, or other power sources. Since many of the scales used a different definition of the satisfaction construct, direct comparisons of these correlations were deemed inappropriate. A subsequent examination of the correlations between power source scales, together with a comparison to the expected direction of the relationship was useful in investigating nomological validity. The results support claims of nomological validity for the measures of reward, referent, and expert power by each author. However, the measure of coercive power used by Comer (1984) and the legitimate power measures used by John (1984) and Hinkin and Schriesheim (1989) do not relate to other power source measures in a predictable fashion.

The measurement of legitimate power is a specific area of concern. First, scale reliabilities for legitimate power were generally lower in comparison to measures of other bases of power. Second, it is noteworthy that the legitimate power scales used by Comer (1984) and the original legitimate power scale used by Kohli (1989) were both found to possess a *bidimensional* structure. Kohli later dropped three of the five items from his scale and re-named it "informal legitimate power" (p. 57). The legitimate power scales used by John (1984) did not correlate in the expected manner with a measure of coercive power. Finally, the legitimate power scale used by Hinkin and Schriesheim (1989) correlated positively ($r = .16$ to $.40$) with a measure of coercive power, when a negative correlation was expected. Additional

research is needed to re-specify and confirm the dimensions of this construct.

Bagozzi and Yi (1991) have suggested and reviewed alternative measures of achieving convergent and discriminant validity, including MTMM and confirmatory factor analysis. Among the power measures reviewed, only Comer (1984) made use of the *full* MTMM approach. A reduced version of the MTMM procedure was used by Gaski (1986, Gaski and Nevin 1985). Only Frazier and Summers (1986), Gaski (1986) and Hinkin and Schriesheim (1989) appear to have used confirmatory factor analysis as a primary means for assessing discriminant validity.

RECOMMENDATIONS

Comer (1984), John (1984), Gaski (1986, Gaski and Nevin 1985), and Hinkin and Schriesheim (1989) each used multi-item scales for all of the five traditional measures of power. Overall, the multi-item scales reported by Gaski (1986, Gaski and Nevin 1985) were found to provide the most reasonable evidence of construct validity. Scale reliabilities were low, but acceptable. Convergent and discriminant validity was claimed and generally supported for each of the power source scales, although the use of the modified MTMM approach and alternative measures which are not maximally different does not provide conclusive support of these claims. The claim of nomological validity for each of the Gaski scales is supported through the use of confirmatory factor analysis, and the correlations between each power source scale are in the direction and magnitude to be

expected. As a set, these scales are considered to be the most reliable and valid measures of the sources of power among those reviewed.

The most complete analysis for establishing construct validity among the scales reviewed was the procedures used by Comer (1984). The reward, expert, and referent scales are supported as being both reliable and valid measures of power. Considerable evidence is provided by Comer to support the reliability and dimensional measurement of these three scales. Scale reliabilities are consistently high, and the procedures used for establishing construct validity are both comprehensive and rigorous. While an overall failure to establish discriminant validity using MTMM was noted by Comer, the source of the failure may be a confounding method variance problem due to the lack of maximally different methods. The inability of MTMM to identify and separate method variance from the data has been previously noted (Bagozzi and Yi 1991). The lack of maximally different methods, however, was a problem in all of the scales reviewed. The convergent and discriminant validity of the scales used by Comer were also assessed using exploratory factor analysis and alternative measures, while nomological validity was examined through correlations of each scale with satisfaction.

Unfortunately, the legitimate and coercive power measures reported by Comer (1984) are not recommended for further use as presently constructed. Correlations of the coercive power measure with referent power measures yielded a positive correlation (a negative one was expected). The correlation

between coercive power and expert power was negligible (a negative correlation was expected.) As previously discussed, the legitimate power measure reported by Comer possessed a bidimensional factor structure.

The power base scales developed and used by Hinkin and Schriesheim (1989) were found to possess high reliabilities, and to have reasonable evidence of discriminant and nomological validity for most of the scales. The results obtained through the use of both exploratory and confirmatory factor analysis generally support the claims for validity of each measure. In particular, the legitimate power measure used by Hinkin and Schriesheim possessed a high reliability coefficient and relatively consistent correlations between the legitimate power scale and other power scales. There is some question, however, regarding the nomological validity of the coercive power measure. Coercive power correlated positively with legitimate power when a negative correlation was expected. In addition, further investigation including the use of alternative methods for each construct are needed before support can be given to the claim of convergent validity.

The power attributions scales reported by John (1984) also possess acceptable reliabilities. While common, the exploratory factor analysis procedure used to assess convergent and discriminant validity does not provide conclusive evidence for the claim of convergent and discriminant validity. A comparison of the correlations between measures provides support for the validity of the reward, referent, and expert power scales.

However, an insignificant correlation was found between the legitimate and coercive power scales, when a negative correlation was expected. Claims of validity for these two scales are inconclusive. The use of these two scales can not be recommended without additional and conclusive validation.

Regarding overall measures of power, the measures used by Gaski and Nevin (1985) and Frazier and Summers (1986) both provide acceptable evidence of reliability and validity. Both supplied support for their claims of convergent and discriminant validity: Gaski and Nevin used a modified version of the MTMM approach to construct validation, while Frazier and Summers used confirmatory factor analysis. A claim of nomological validity is particularly well supported for both scales, as both scales used correlations with multiple theoretically related constructs.

CONCLUSION

The establishment of reliable and valid measures of power sources is a difficult but important undertaking. The development and testing of many hypotheses in the areas of sales and channels involving power is contingent upon measures of the power construct which are reliable and valid. Existing research has successfully established several multi-item measures that possess satisfactory reliability coefficients. However, increased attention is needed toward establishing measures of power and power sources which have convergent, discriminant, and nomological validity.

Increased attention is needed for the development of better measures of the legitimate power construct. Legitimate power was found to have a generally low reliability coefficient and weaker evidence of nomological validity (in comparison to other power scale measures). Of the power scales reviewed, the power source scales used by Gaski (1986, Gaski and Nevin 1985) and the overall measures of power used by Gaski and Nevin (1985) and Frazier and Summers (1986) were found to be both generally reliable and valid, and are recommended for further use.

****

TABLE 1
COMPARISON OF SCALE CHARACTERISTICS

****

Author and Year	Scale Name	Sample	Scale Items	Scale Pts.	Coeff. Alpha
Reward Power Scales:					
John (1984)	Reward Attributions	147	6	5	0.92
Gaski & Nevin (1985)	Exercised Reward Power	238	15	4	0.83
Gaski & Nevin (1985)	Reward Power Source	238	15	5	0.87
Comer (1984)	Reward Power Source	207	5	7	0.88
Hinkin & Schriesheim (1989)[1]	Reward Power	251	4	5	0.80
		375	*	*	0.77
		220	*	*	0.77
Means			9	5.2	0.86
Coercive Power Scales:					
John (1984)	Coercive Attributions	147	4	5	0.90
Gaski & Nevin (1985)	Exercised Coercive Power	238	6	4	0.62
Comer (1984)	Coercive Power Source	207	5	7	0.74
Gaski & Nevin (1985)	Coercive Power Source	238	6	5	0.69
Michie & Sibley (1985)	Coercive Power Source	123	8	5	Not reported
Hinkin & Schriesheim (1989)[1]	Coercive Power	251	4	5	0.80
		375	*	*	0.85
		220	*	*	0.86
Means			5.5	5.17	0.76
Referent Power Scales:					
John (1984)	Referent Attributions	147	4	5	0.83
Gaski (1986)	Referent Power Source	238	5	5	0.81
Kohli (1989)	Referent Power Source	251	5	5	0.86
Comer (1984)	Referent Power Source	207	5	7	0.87
Hinkin & Schriesheim (1989)[1]	Referent Power	251	4	5	0.88
		375	*	*	0.87
		220	*	*	0.86
Means			4.6	5.4	0.85
Expert Power Scales:					
John (1984)	Expert Attributions	147	4	5	0.79
Gaski (1986)	Expert Power Sources	238	5	5	0.77
Kohli (1989)	Expert Power Sources	251	4	5	0.85
Comer (1984)	Expert Power Sources	207	5	7	0.89
Hinkin & Schriesheim (1989)[1]	Expert Power	251	4	5	0.85
		375	*	*	0.90
		220	*	*	0.83
Means			4.4	5.4	0.83
Legitimate Power Scales:					
John (1984)	Legitimate Attributions	147	6	5	0.69
Comer (1984)	Legitimate Power Sources	207	5	7	0.68
Gaski (1986)	Legitimate Power Sources	238	5	5	0.65
Hinkin & Schriesheim (1989)[1]	Legitimate Power	251	4	5	0.87
		375	*	*	0.85
		220	*	*	0.86
Means			5	5.5	0.72
Power:					
Gaski & Nevin (1985)	Power	238	10	4	0.87
Kohli (1989)	Departmental Power	251	4	5	0.88
Butaney & Wortzel (1988)	Distributor Power	83	17	5	0.85
Frazier & Summers (1986)	Manufacturer Power	435	6	11	0.81
Means			9.25	6.25	0.84

[1] = Hinkin and Schriesheim (1989) used each scale (reward, coercive, etc.) with three different samples. Sample sizes and coefficient alphas are reported for each scale and sample. To compute the means, the Hinkin and Schriesheim scales were counted once; coefficient alphas for the three samples were averaged, with the average for the three samples used for computing the mean for the studies.

****

TABLE 2

CORRELATIONS BETWEEN CRONBACH'S ALPHA AND
SAMPLE SIZE, NUMBER OF SCALE ITEMS, AND NUMBER OF SCALE POINTS

****

	Correlation with Cronbach's Alpha	p-value
Sample Size	0.08150	0.6267
Number of Scale Items	0.03790	0.8213
Number of Scale Points	-0.04350	0.7954

****

****

TABLE 3

COMPARISON OF CONVERGENT AND DISCRIMINANT VALIDITY PROCEDURES

****

Author & Year	Scale Name	Method of Assessing Convergent and Discriminant Validity
Reward Power Scales:		
John (1984)	Reward Attributions	Exploratory Factor Analysis
Gaski and Nevin (1985)	Exercised Reward Power	MTMM[1], Alternative Measures
Gaski and Nevin (1985)	Reward Power Source	MTMM[1], Alternative Measures
Comer (1984)	Reward Power Source	MTMM, Alternative Measures, Exploratory Factor Analysis
Hinkin and Schriesheim (1989)	Reward Power	Exploratory Factor Analysis, Confirmatory Factor Analysis
Coercive Power Scales:		
John (1984)	Coercive Attributions	Exploratory Factor Analysis
Gaski and Nevin (1985)	Exercised Coercive Power	MTMM[1], Alternative Measures
Comer (1984)	Coercive Power Source	MTMM, Alternative Measures, Exploratory Factor Analysis
Gaski and Nevin (1985)	Coercive Power Source	MTMM[1], Alternative Measures
Michie and Sibley (1985)	Coercive Power Source	Exploratory Factor Analysis, Alternative Measures
Hinkin and Schriesheim (1989)	Coercive Power	Exploratory Factor Analysis, Confirmatory Factor Analysis
Referent Power Scales:		
John (1984)	Referent Attributions	Exploratory Factor Analysis
Gaski (1986)	Referent Power Source	MTMM[1]
Kohli (1989)	Referent Power Source	Exploratory Factor Analysis
Comer (1984)	Referent Power Source	MTMM, Alternative Measures, Exploratory Factor Analysis
Hinkin and Schriesheim (1989)	Referent Power	Exploratory Factor Analysis, Confirmatory Factor Analysis
Expert Power Scales:		
John (1984)	Expert Attributions	Exploratory Factor Analysis
Gaski (1986)	Expert Power Sources	MTMM[1]
Kohli (1989)	Expert Power Sources	Exploratory Factor Analysis
Comer (1984)	Expert Power Sources	MTMM, Alternative Measures, Exploratory Factor Analysis
Hinkin and Schriesheim (1989)	Expert Power	Exploratory Factor Analysis, Confirmatory Factor Analysis
Legitimate Power Scales:		
John (1984)	Legitimate Attributions	Exploratory Factor Analysis
Comer (1984)	Legitimate Power Sources	MTMM, Alternative Measures[2]
Gaski (1986)	Legitimate Power Sources	MTMM[1]
Hinkin and Schriesheim (1989)	Legitimate Power	Exploratory Factor Analysis, Confirmatory Factor Analysis
Power:		
Gaski & Nevin (1985)	Power	MTMM[1]
Kohli (1989)	Departmental Power	Exploratory Factor Analysis
Butaney & Wortzel (1988)	Distributor Power	Exploratory Factor Analysis
Frazier & Summers (1986)	Manufacturer Power	Confirmatory Factor Analysis

[1] = refers to a *modified* version of the multitrait-multimethod approach.
[2] = Convergent and discriminant validity were reported by Comer (1984) as not established. Legitimate power was found to have a bidimensional factor structure.

****

TABLE 4

COMPARISON OF NOMOLOGICAL VALIDITY PROCEDURES

****

Author & Year	Scale Name	Method of Assessing Nomological Validity
Reward Power Scales:		
John (1984)	Reward Attributions	Correlation with other power sources, attitude components, formalization and centralization
Gaski & Nevin (1985)	Exercised Reward Power	Correlation with satisfaction
Gaski & Nevin (1985)	Reward Power Source	Correlation with satisfaction
Comer (1984)	Reward Power Source	Correlation with satisfaction
Hinkin & Schriesheim (1989)	Reward Power	Correlations with three measures of satisfaction
Coercive Power Scales:		
John (1984)	Coercive Attributions	Correlation with other power sources, attitude components, formalization and centralization
Gaski & Nevin (1985)	Exercised Coercive Power	Correlation with satisfaction
Comer (1984)	Coercive Power Source	Correlation with satisfaction
Gaski & Nevin (1985)	Coercive Power Source	Correlation with satisfaction
Michie & Sibley (1985)	Coercive Power Source	Not specified
Hinkin & Schriesheim (1989)	Reward Power	Correlations with three measures of satisfaction
Referent Power Scales:		
John (1984)	Referent Attributions	Correlation with other power sources, attitude components, formalization and centralization
Gaski (1986)	Referent Power Source	Correlation with power, power sources, satisfaction
Kohli (1989)	Referent Power Source	Not specified
Comer (1984)	Referent Power Source	Correlation with satisfaction
Hinkin & Schriesheim (1989)	Referent Power	Correlations with three measures of satisfaction
Expert Power Scales:		
John (1984)	Expert Attributions	Correlation with other power sources, attitude components, formalization and centralization
Gaski (1986)	Expert Power Sources	Correlation with power, power sources, satisfaction
Kohli (1989)	Expert Power Sources	Not specified
Comer (1984)	Expert Power Sources	Correlation with satisfaction
Hinkin & Schriesheim (1989)	Expert Power	Correlations with three measures of satisfaction
Legitimate Power Scales:		
John (1984)	Legitimate Attributions	Correlation with other power sources, attitude components, formalization, and centralization
Comer (1984)	Legitimate Power Sources	Correlation with satisfaction
Gaski (1986)	Legitimate Power Sources	Correlation with power, power sources, satisfaction
Hinkin & Schriesheim (1989)	Legitimate Power	Correlations with three measures of satisfaction

****

TABLE 4
COMPARISON OF NOMOLOGICAL VALIDITY PROCEDURES
(continued)

****

Power:		
Gaski & Nevin (1985)	Power	Correlation with expert, referent, and legitimate power sources
Kohli (1989)	Departmental Power	Not specified
Butaney & Wortzel (1988)	Distributor Power	Not specified
Frazier & Summers (1986)	Manufacturer Power	Correlation with satisfaction, interest of the source firm in the target firm's welfare, and agreement on general business strategy

****

TABLE 5

CORRELATIONS BETWEEN POWER MEASURES

****

Author and Year	Scale Name	Correlations w/ Other Power Scales				
Reward Power Scales:		Reward	Coercive	Referent	Expert	Legit.
Comer (1984)	Reward Power Source	----	.22	.61	.57	.31 John
(1984)	Reward Attributions	----	.39	.20	.20	.39
Gaski & Nevin (1985)	Exer. Reward Power	----	-.17	.47	.47	.23
Gaski & Nevin (1985)	Reward Power Source	----	.27	----	----	----
Hinkin & Schriesheim (1989)[1]	Reward Power	----	.06	.35	.27	.29
		----	.11	.40	.43	.36
		----	.14	.32	.29	.29
Expected Direction of Correlation:		----	+	+	+	+
Coercive Power Scales:						
Comer (1984)	Coerc. Power Source	.22	----	.25	.02	-.12
John (1984)	Coerc. Attributions	.39	----	-.28	-.27	.01
Gaski & Nevin (1985)	Exer. Coercive Pwr.	-.26	----	-.37	-.42	-.24
Gaski & Nevin (1985)	Coerc. Power Source	.27	----	----	----	----
Hinkin & Schriesheim (1989)[1]	Coercive Power	.06	----	-.07	-.06	.20
		.11	----	-.06	-.08	.16
		.14	----	.16	.06	.41
Expected Direction of Correlation:		+	----	-	-	-
Referent Power Scales:						
Comer (1984)	Refer. Power Source	.73	.25	----	.61	.24
John (1984)	Refer. Attributions	.20	-.28	----	.70	.45
Gaski (1986)	Refer. Power Source	.47	-.37	----	.69	.49
Hinkin & Schriesheim (1989)[1]	Referent Power	.35	-.07	----	.52	.53
		.40	-.06	----	.72	.57
		.32	.16	----	.52	.55
Expected Direction of Correlation:		+	-	----	++	++
Expert Power Scales:						
Comer (1984)	Expert Power Sources	.57	.02	.61	----	.41
John (1984)	Expert Attributions	.20	-.27	.70	----	.50
Gaski (1986)	Expert Power Sources	.47	-.42	.69	----	.41
Hinkin & Schriesheim (1989)[1]	Expert Power	.27	-.06	.52	----	.46
		.43	-.08	.72	----	.64
		.29	.06	.52	----	.39
Expected Direction Correlation:		+	-	++	----	++
Legitimate Power Scales:						
Comer (1984)	Legit. Power Sources	.31	-.12	.24	.41	----
John (1984)	Legit. Attributions	.39	.01	.45	.50	----
Gaski (1986)	Legit. Power Sources	.23	-.24	.49	.41	----
Hinkin & Schriesheim (1989)[1]	Legitimate Power	.29	.20	.53	.46	----
		.36	.16	.57	.64	----
		.29	.41	.55	.39	----
Expected Direction of Correlation:		+	-	++	++	----

Notes:
+ = a modest and positive correlation is expected
++ = a strong and positive correlation is expected
- = a negative correlation is expected

Footnote:
[1] = Hinkin and Schriesheim (1989) used three separate samples. Sample "A" consisted of 251 upper-level undergraduate business students from a southern university. Sample "B" consisted of 375 full-time employees of a large southern psychiatric hospital. Sample "C" consisted of

220 part-time MBA students from a large southern university. Correlation coefficients for all three samples are reported, in order (Sample "A" first, etc.).

Note: Correlation coefficients between power scales were not reported by Kohli (1989) or Michie and Sibley (1985).

****

References

Bagozzi, Richard P. and Lynn W. Phillips (1982), "Representing and Testing Organizational Theories: A Holistic Construal," Administrative Science Quarterly, 27 (September), 459-489.

_____ and Youjae Yi (1991), "Multitrait-Multimethod Matrices in Consumer Research," Journal of Consumer Research, 17 (March), 426-439.

Burke, R. J. and Wilcox, D. S. (1971), "Bases of Supervisory Power and Subordinate Job Satisfaction," Canadian Journal of Behavioral Sciences, 3, 183-193.

Busch, Paul (1980), "The Sales Manager's Bases of Social Power and Influence on the Sales Force," Journal of Marketing, 44 (Summer).

Butaney, Gul and Lawrence H. Wortzel (1988), "Distributor Power Versus Manufacturer Power: The Customer Role," Journal of Marketing, 52 (January), 52-63.

Campbell, Donald R. and Donald W. Fiske (1959), "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," Psychological Bulletin, 56, 81-105.

Churchill, Gilbert A. Jr. (1976), Marketing Research: Methodological Foundations. Hinsdale, IL: Dryden.

_____ (1979), "A Paradigm for Developing Better Measures of Marketing Constructs," Journal of Marketing Research, 16 (February), 64-75.

_____ and J. Paul Peter (1984), "Research Design Effects on the Reliability of Rating Scales: A Meta-Analysis," Journal of Marketing Research, 21 (November), 360-375.

Cohen, Joel B. (1979), "Exploring Attitude Construct Validity: Or Are We?" in Advances in Consumer Research Volume 6, W. L. Wilkie, ed. Ann Arbor: Association for Consumer Research.

Comer, James M. (1984), "A Psychometric Assessment of a Measure of Sales Representatives' Power Perceptions," Journal of Marketing Research, 21 (May) 221-225.

Cronbach, Lee J. (1971), "Test Validation," in Educational Measurement Second Edition, R. L. Thorndike, ed. Washington, D.C.: American Council on Education.

DeVellis, Robert F. (1991), Scale Development: Theory and Applications. Newbury Park: Sage.

El-Ansary, Adel I. and Louis W. Stern (1972), "Power Measurement in the Distribution Channel" in Contemporary Issues in Marketing Channels, Robert F. Lusch and Paul H. Zinszer, eds. Norman, OK: University of Oklahoma.

Etgar, Michael (1976), "Channel Domination and Countervailing Power in Distributive Channels," Journal of Marketing Research, 13 (August) 254-262.

_____ (1978), "Interchannel Conflict and the Use of Power." Journal of Marketing Research, 15 (May), 273-274.

Frazier, Gary L. (1983), "On the Measurement of Interfirm Power in Channels of Distribution," Journal of Marketing Research, 20 (May) 158-166.

_____ and John O. Summers (1986), "Perceptions of Interfirm Power and Its Use Within a Franchise Channel of Distribution," Journal of Marketing Research, 23 (May), 169-176.

French, John R. P. and Bertram H. Raven (1959), "The Bases of Social Power," in Studies in Social Power, Dorwin Cartwright, ed. Ann Arbor: University of Michigan Press.

Gaski, John F. (1986), "Interrelations Among a Channel Entity's Power Sources: Impact on the Exercise of Reward and Coercion on Expert, Referent, and Legitimate Power Sources," Journal of Marketing Research, 23 (February), 62-77.

_____ and John R. Nevin (1985), "The Differential Effects of Exercised and Unexercised Power Sources in a Marketing Channel," Journal of Marketing Research, 22 (May), 130-142.

Gerbing, David W. and James C. Anderson (1988), "An Updated Paradigm for Scale Development Incorporating Unidimensionality and Its Assessment," Journal of Marketing Research, 25 (May), 186-192.

Hinkin, Timothy R. and Chester A. Schriesheim (1989), "Development and Application of New Scales to Measure the French and Raven (1959) Bases of Social Power," Journal of Applied Psychology, ? (4) 561-567.

Holzbach, Robert Lawrence Jr. (1974), "An Investigation of a Model for Managerial Effectiveness: The Effects of Leadership Style and Leader Attributed Social Power on Subordinate Job Performance" Dissertation, Carnegie-Mellon University.

Hunt, Shelby D. and John R. Nevin (1974), "Power in a Channel of Distribution: Sources and Consequences," Journal of Marketing Research, 11 (May), 186-193.

John, George (1984), "An Empirical Investigation of Some Antecedents of Opportunism in a Marketing Channel," Journal of Marketing Research, 21 (August), 278-289.

Joreskog, Karl G. and Dag Sorbom (1984), LISREL VI: Analysis of Linear Structural Relationships by the Method of Maximum Likelihood. Chicago: National Educational Resources.

Kohli, Ajay (1989), "Determinants of Influence in Organizational Buying: A Contingency Approach," Journal of Marketing, 53 (July), 50-65.

Lusch, Robert F. (1977), "Franchisee Satisfaction: Causes and Consequences," International Journal of Physical Distribution, 7, 128-140.

Michie, Donald A. and Stanley D. Sibley (1985), "Channel Member Satisfaction: Controversy Resolved," Journal of the Academy of Marketing Science, 13 (Spring), 188-205.

Nunnally, Jum (1978), Psychometric Theory, Second Edition. New York: McGraw Hill.

Peter, J. Paul (1979), "Reliability: A Review of Psychometric Basics and Recent Marketing Practices," Journal of Marketing Research, 16 (February) 6-17.

_____ (1981), "Construct Validity: A Review of Basic Issues and Marketing Practices," Journal of Marketing Research, 18 (May), 133-145.

Podsakoff, Philip M. and Chester A. Schriesheim (1985), "Field Studies of French and Raven's Bases of Power: Critique, Reanalysis, and Suggestions for Future Research," Psychological Bulletin, 97 (3) 387-411.

Raven, Bertram H. (1965), "Social Influence and Power," in Current Issues in Social Psychology, Ivan D. Steiner and Martin Fishbein, eds. New York: Holt, Rinehart, and Winston.

_____ and Arie W. Kruglanski (1970), "Conflict and Power," in The Structure of Conflict, Paul Swingle, ed. New York: Academic Press.

Schul, Patrick L., Steven J. Remington, and Robert L. Berl (1990), "Assessing Gender Differences in Relationships Between Supervisory Behaviors and Job-Related Outcomes in

the Industrial Sales Force," Journal of Personal Selling and Sales Management, 10 (Summer) 1-16.