Scale Development for E-Procurement (EP) 
Adoption Influence Tactics

Marc R.B. Reunis
Delft University of Technology, The Netherlands

Erik M. van Raaij
Eindhoven University of Technology, The Netherlands

Summary

In this paper, scales are developed for measuring proactive influence tactics which can be used to encourage individual adoption of electronic procurement (EP). Different perspectives on scale development are combined into one scale development approach, and this approach is used to build an instrument for measuring the utilization of seven influence tactics. An item battery is generated and initial scales are evaluated with respect to their psychometric properties. This paper intends to make both a methodological and theoretical contribution. It initiates the discussion for improved scale development in purchasing research and elaborates our understanding of influence tactics in EP adoption.

Keywords

Scale Development, Influence Tactics, E-Procurement

Educator and practitioner summary

In order to reap full benefits from the implementation of e-procurement tools, these tools need to be adopted and used by the intended users in an organisation. Managers can use a variety of influence tactics to stimulate such user adoption. In this paper, the authors develop scales with which the incidence and effectiveness of seven influence tactics can be measured, with the ultimate aim to be able to help managers decide which influence tactics to use in future EP implementations.

Introduction

Over the past decade, a wide variety of electronic procurement (EP) tools have changed and continue to change the purchasing landscape. The potential benefits of EP applications, such as reverse auctions, purchase-to-pay and supplier relationship management systems, have been widely recognized in practice, showing shorter order cycle times, cost saving opportunities, the possibility to leverage spend, and create compliance to processes and contracts (De Boer et al., 2002). Attractive business cases, however, can ‘evaporate’ when adoption by end-users lags

1 PhD Candidate at the Department of Aerospace Management & Operations (AMO), Faculty of Aerospace Engineering, Delft University of Technology (TUD), P.O. Box 5058, 2629 HS Delft, The Netherlands, t: +31(0)151781483, f: +31(0)152783444, m.r.b.reunis@lr.tudelft.nl (corresponding author)
2 Assistant Professor at the Department of Technology Management, Eindhoven University of Technology, P.O. Box 516, 5300 MB Eindhoven, The Netherlands, t: +31(0)402472629, f: +31(0)402468054, e.m.vanraaij@tm.tue.nl
behind. The benefits can only be achieved if and when the end-users adopt the tool and apply it successfully in their every day work, i.e. use the system in the ‘right’ way and to its full potential. End-user adoption is not self-evident, however. People are not keen on altering their habits and can generally be reluctant to change. Resistance to change can be particularly salient in an intra-organizational setting, where the tool is implemented top-down and where an apparent benefit may not be readily perceived at the level of the individual user. Individuals tend to minimize their effort to comply when they are not committed to the change (Reunis, Santema, 2005). Currently, it is not clear how to speed up and/or raise the overall level of intra-organizational adoption. The number of studies focusing on EP adoption is growing rapidly, but most of these studies focus on the tool, and not on the user. Individual adoption of these systems has only briefly been addressed (e.g. Reunis, 2005). Change management principles can provide some guidance, but practitioners often do not know what/which influence tactic to use, and when. Based on both an extensive review of the social behavioural psychology literature and on our own studies of EP roll-outs, seven influence tactics have been identified, that managers may use in their efforts to stimulate user adoption of EP tools. These seven tactics are: persuasive recommendation (PREC), management request (MR), information push (IP), collaboration (COLL), legitimate pressure (LEGP), reward (REW), and appraisal & control (A&C). The definitions of these seven influence tactics are new and existing measurement scales for these influence tactics are not readily available. Such measurement scales are needed in order to evaluate the use and effectiveness of these influence tactics (cf. Reunis et al., 2004).

The importance of appropriate scales for measuring constructs is paramount and has been pointed out by many researchers from different disciplines within social science (Churchill, 1979; Hinkin, 1995; Rossiter, 2002; Straub et al., 2004). In order to draw any kind of conclusion regarding relations between latent constructs, scales must be constructed so as to capture the ‘true’ concept. The observation of true covariance between the variables of interest is directly dependent upon the ability to accurately and reliably operationalise the unobservable construct in a scale. To ensure adherence to criteria of quality, specific approaches have been suggested for the process of scale construction. Some of these approaches emphasize the conceptual quality of the scale (e.g., Rossiter, 2002) and others the scale's psychometric properties (e.g., Churchill, 1979).

A quick scan of papers published in JPSM, JSCM, and the proceedings of IPSERA and IMP shows that within purchasing, only limited explicit attention has been given to processes of scale development. From the 190 articles with keywords "scale", "measure" and "measurement", only 14 articles develop a measurement instrument for latent constructs (e.g., relationship atmosphere, e-readiness, inter-firm trust, risk management, cross-functional orientation, relationship value, communication capabilities, buyer-supplier relationships). From these 14 articles, 8 use a sample for assessing psychometric properties; generally with factor analysis and reliability. None of them explicitly discuss the conceptual quality of the scale. Overall, only a few articles provide a step-by-step description of the process that was followed (e.g. Ulaga, Eggert, 2001; Large, Giménez, 2005; Eng, 2005). None of them include a critical view of the process itself.

In the marketing discipline, however, discussions regarding scale development are more widespread, and can be traced back to the seminal article of Churchill (1979). Churchill (1979) calls for better measures for marketing constructs, and proposes an eight-step method with an emphasis on obtaining favourable psychometric properties. The Churchill approach has become
the dominant paradigm in marketing and has improved scale construction by providing a step-by-step approach for conceptualisation and validation. Rossiter (2002) challenged the Churchill approach and proposed a six-step procedure called C-OAR-SE, which is an acronym for construct definition (C), object classification (O), attribute classification (A), rater identification (R), scale formation (S) and enumeration and reporting (E). This approach emphasizes the conceptualisation of constructs, thereby addressing a weakness in the Churchill paradigm. Finn & Kayande (2005) support Rossiter's call for more attention for conceptualisation and content validity, but criticize the extreme context dependence. While Churchill seems to overemphasize the empirical validation using respondent samples at the cost of conceptual rigor, Rossiter seems to overemphasize the contextual nature. Following the criticism of Finn & Kayande (2005), we strike a balance between the two approaches and propose a three-step approach for scale development for latent reflective constructs, based on two reviews of best practices in operations management (Hensley, 1999) and marketing (Hinkin, 1995):

1. **Item generation.** The generation of items may be the most important part of developing sound measures. Content validity is the main aim of this step, which can be accomplished by a strong guiding theoretical framework and employing a rigorous sorting process that matches items to construct definitions. Both inductive and deductive approaches can be distinguished. Here, a combined approach is suggested.

2. **Scale development**
   a. **Design of study.** The aim of this step is to examine the psychometric properties of a new measure. Choosing the specific sample, wording of items, reducing pattern response bias, and scale length are important issues to consider here.
   b. **Scale construction.** Factor analysis techniques are used to assess the stability of the factor structure (dimensionality) and provide information that will facilitate the refinement of the new measure. The elimination of poorly loading items is an iterative balancing act between psychometric properties and content validity. In this stage, poor item development practices surface: scales should not be derived post hoc.
   c. **Reliability assessment.** In this step, consistency of items within a measure and the stability of the measure over time is assessed.

3. **Scale Evaluation.** A new sample is used to evaluate the psychometric properties of the scale: construct validity (discriminant and convergent validity), as well as criterion-related validity.

This paper aims to make both a theoretical and a methodological contribution to the field of purchasing. It initiates the discussion for improvements in scale development in purchasing research and enhances our understanding of influence tactics in EP adoption. The three-step approach, balancing Churchill’s and Rossiter’s view, is applied to develop measures for the seven influence tactics. First of all, the theoretical background of influence tactics is described and a need for new scales is established. Next, an item battery is generated based upon both previous power and influence theories (deductive) and logical partitioning (inductive). The items are combined on the basis of inter-rater reliability. This concludes step 1. Thereafter, the initial scales are evaluated with respect to their psychometric properties using a small empirical sample. The study design, factor analysis and reliability assessment are established This concludes step 2. In the subsequent section, the process of balancing both content and construct validity is described, and the final scale is proposed for a test in step 3. Finally, the process of scale development is discussed.
Theoretical background on Influence Tactics

The concepts of influence tactics are closely related to social power. Both concepts are grounded in early work in social psychology (Bruins, 1999). A seminal work in the field of power and influence is the paper of French & Raven (1959), who defined influence as a force one person (the agent) exerts on someone else (the target) to induce a change in the target, including changes in behaviour, opinion, attitude, goals, needs, and values. Social power was subsequently defined as the potential ability of an agent to influence a target. Thus, influence is "kinetic power, just as power is potential influence" (French, Raven, 1959, p. 152). Power theories have been used extensively in relation to the choice and application of certain influence tactics or measures. Power can thereby be seen as a potential which is effectuated by certain influences. On the basis of perceived power, which, in turn, is dependent on resource availability and dependency, an agent can choose between a variety of influence tactics to reach a desired change.

Classifying Influence Tactics

Research on influence tactics has developed in two distinct ways. The first stream of research builds on strong theoretical grounds of power theories and can be characterized as deductive research. Empirical work in this stream mainly focuses on finding support for theoretical concepts on social power and social influence. The second research stream takes a more inductive approach towards classifying influence tactics on the basis of empirical findings. In this type of research, ways in which people attempt to influence each other are identified, clustered and subsequently explained in theoretical terms. The advantage of the first approach is scientific rigor. The question remains, however, if the theory is exhaustive (i.e., whether it covers all influence tactics people use). Conversely, the main advantage of the second approach is the completeness, given a certain context, and the empirical description obtained in this way. The underlying mechanisms of causal relations, i.e. the process of influence, of the second approach is weaker.

The main proponents of the deductive stream are French & Raven (1959), who proposed a classification with five different bases of power: reward, coercive, legitimate, referent, and expert power. A sixth basis (informational power) was added later on (Raven, 1965). *Reward* and *coercive* power refer to the agent's ability to generate positive and negative outcomes with respect to the target. Using either of these bases will induce only a superficial change in the target, leading to compliance. The continuation depends upon control of the target by the agent. *Legitimate* power is based on the target's belief that the agent has a legitimate right to exert influence, and that the target has an obligation to accept this influence. *Referent* power depends on the target's identification with the agent. This power base also leads to acceptance by the target as it enables the target to maintain a relationship with the agent or to identify with the agent. *Expert* power is the attribution of superior knowledge or experience to the agent. This power basis leads to acceptance. *Informational* power is based on access and control over information. Information from various sources can lead to internalised and lasting changes in the target's beliefs, attitudes or values. An agent can influence this process by controlling the information flow. This informational power is independent of the person and the agent's relationship with the target. The perceived relevance and validity of the information is of utmost importance. The original and revised classification of power bases remains an important theoretical basis for deductive empirical work on influence tactics.
Various inductive classifications or taxonomies have been compiled since the late 1970s. A well-known example is Kipnis (1976), who studied influence processes inductively from the perspective of the agent. Kipnis (1976) concluded that the choice of influence tactics depends upon a rational decision process from the part of the agent, taking into account the resources available (power basis), the inhibitions of using a power basis, and on the resistance expected by the agent. As a follow-up, Kipnis et al. (1980) investigated the "means of influence" by identifying the specific behaviours people have at their disposal to influence others. Kipnis et al. (1980) found eight categories of tactics: assertiveness, ingratiation, rationality, sanctions, exchange, upward appeal, blocking, and coalitions. A similar classification has been developed by Yukl and his colleagues (Yukl, Falbe, 1990; Yukl, Tracey, 1992). Many debates on how to identify, define and cluster influence tactics have ensued among social psychologists and communication scholars, resulting in a plethora of classifications, underlying dimensions and clusters. Work on social influence since the 1950s has resulted in at least 22 different item batteries for measuring susceptibility, influence strategies, and power of influence.

Influence Tactics for EP

The generalization of existing scales to new, unexplored contexts, is an issue which is raised by Rossiter (2002). In his view, every application of a scale is to some extent idiosyncratic. Rossiter (2002) explicitly recognizes the context specificity of a scale, and calls for particular attention to object, attributes and rater in the process of construct definition and item development. Rossiter (2002) opposes the idea that in purely positivistic studies, which aim to identify general governing ‘laws’, scales are generally believed to be the same across contexts and applications. In our research, EP is the object and potential end-users of EP tools are the raters. The direct applicability of previously used scales for generic influence tactics is limited in our study for the following two reasons:

- We cannot assume without additional research whether existing general scales cover the incidence of different types of influence tactics in the specific context of EP adoption.
- General scales of influence tactics are not balanced for the occurrence of influence tactics for EP adoption: some tactics occur frequently, while other do not occur at all. For the purpose of post-hoc analysis of the effectiveness of influence tactics, they must have occurred in practice before their effectiveness can be assessed using a scale. For instance, Venkatesh et al. (1995) include ‘threats’ in their influence tactics scale. Yet, four exploratory case studies on EP adoption show no incidence of this tactic.

Therefore, the specific context of this research calls for specific scales. The scales developed in this paper should at least cover the influence tactics which are relevant for EP adoption and not include redundant or extraneous ones. Obviously, some items from previously developed general scales can be used. This is assessed in the next section.

Step 1: Item Generation

A combined deductive and inductive approach was followed to develop a battery of measurement items. The prime focus of item generation is establishing content validity of the latent constructs. The measures must capture the intended content through the items in the scale. This means that a.) an item battery should be developed and b.) items should be selected from the item battery to cover the concept yet contain no extraneous content. Regrettably, there is no quantitative or qualitative measure that tests whether the scale adequately captures the specific
domain of interest. Expert judgement is therefore of utmost importance in assessing the content validity of a measure (Stone, 1978).

The deductive approach takes a firmly grounded theoretical definition of the concept under scrutiny as a guide for the development of items. Existing classification schemes or typologies from prior research can be used to develop new items or select existing ones from previous research. An in-depth understanding of the investigated phenomenon is required for this approach. Researchers can either derive items from previously defined theoretical concepts or develop conceptual definitions grounded in theory, and then utilize a sample of experts to generate items (e.g. by using a critical incidents technique). Hinkin (1995) refers to the deductive approach as "logical partitioning" or "classification from above." The deductive approach fits more in a Churchill view of scale development.

The inductive approach starts from empirical reality instead of established theory and identifies or defines constructs based on what individuals in a specific context say and do, using descriptions, narratives or observations. These are then analysed with respect to content and clustered into a number of categories. Next, labels and definitions are added based upon the empirical findings. Hinkin (1995) calls the inductive approach “classification from below”. This approach fits more with Rossiter’s C-OAR-SE method.

Item battery generation

The deductive and inductive approaches for item battery generation were combined by relaxing the emphasis on previously used definitions of constructs, and rely more on the input from respondents so as to specify them within the context in which they are used. The theoretical classification of power bases by French & Raven was used as a starting point for two case studies in which EP is rolled-out in a multi-national organization. A total of 40 interviews were conducted (14 in the first case, 26 in the second). The interviews were structured according to the six influence tactics based on French & Raven’s power bases and individuals were requested to discuss the manner by which they were influenced (by whom, which activities, which effect). Analysis of the interview data resulted in a tentative definition of seven tactics: persuasive recommendation (PREC), management request (MR), information push (IP), collaboration (COLL), legitimate pressure (LEGP), reward (REW), and appraisal & control (A&C). These definitions of seven influence tactics are new and no existing scales are readily available. Content analysis of the interview responses yielded wording for items for the seven tactics. A total of 35 new items were induced from the field data.

A review of existing items served as an additional input for the item battery. Previous scales were reviewed on item level, not on construct level. Previously used scales for influence tactics were identified through a key-word search in the ETS Database of Psychological measures. A long list of 94 scales was identified using ‘influence’ as keyword. After initial review, 22 scales were deemed relevant for this study. These were subsequently reviewed more carefully for their relevance and 7 scales remained. Of these scales, 69 items were added to the item battery long list, resulting in a total item battery of 104 questions from both previous research and case interviews. The labels of the seven influence tactics as well as their definitions are based on their

---


4 The ETS Database of Psychological Scales is a recommended database of the APA and contains records on over 10,000 tests and research instruments. See: http://sydneyplus.ets.org/
respective theoretical origins, refined and revised on the basis of the case studies. The definitions are still tentative at this point. They are refined during a sorting process, described below.

**Item selection**

After determining an item battery, or long list of potential questions to adequately measure the latent constructs, all items are subjected to a sorting process. This process serves as a pretest, permitting the deletion of items that are deemed to be conceptually inconsistent (Hinkin, 1995). A convenience sample of 19 respondents was used to cluster the 104 items into the seven predefined categories. The clustering is essentially an analytical process that does not require deep knowledge of the phenomena under investigation and can therefore rely on input from MSc / PhD students with good English language skills. The clustering effort provides insight into the relationships of the items with the proposed constructs. Inter-rater reliability was used as a proxy measure for the loading items on constructs. Based on a combination of a.) high inter-rater reliability, b.) conceptual completeness of the construct and c.) expert discussions on content validity, the long-list of 104 items was condensed into a short list of 41 items (6 items for 6 constructs, 5 for one remaining construct). The number of items per construct is a delicate balance: more items could contribute towards a more reliable measure, however, increasing survey length could lead to lower response rates or unreliable data as a result of annoyance on the part of the respondent. Five or six items per construct is believed to balance these issues (Hinkin, 1995). Another issue in constructing the short list is the fit with the theoretical concepts. As mentioned before, no strict theoretical definition is used to guide the sorting process, but the sorting process itself contributed towards the conceptual definition. The list of 41 items formed the input for the next step. Table 1 lists the seven influence tactics and summarizes the process of item generation.

**Table 1  Item battery for the Influence Tactics**

<table>
<thead>
<tr>
<th>Influence Tactic Constructs</th>
<th>Abbr.</th>
<th>Conceptual definition</th>
<th>From theory</th>
<th>From cases</th>
<th>In long list</th>
<th>After sorting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Management Request</td>
<td>MR</td>
<td>A manager who asks to use the system or accept his/her ideas / suggestions, states his/her wishes, and/or asks to comply to his/her request to influence the target individual</td>
<td>6</td>
<td>4</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>Information Push</td>
<td>IP</td>
<td>A source that provides factual information about characteristics of the system, the workings of systems, its performance, and how to operate it.</td>
<td>9</td>
<td>7</td>
<td>16</td>
<td>6</td>
</tr>
<tr>
<td>Persuasive Recommendation</td>
<td>PREC</td>
<td>A source that uses convincing argument and business effects to influence, convert and counter hesitations of the target individual</td>
<td>14</td>
<td>5</td>
<td>19</td>
<td>6</td>
</tr>
<tr>
<td>Collaboration</td>
<td>COLL</td>
<td>Joint effort with an expert or manager to determine system suitability and specification to influence the target individual</td>
<td>5</td>
<td>7</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>Legitimate Pressure</td>
<td>LEGP</td>
<td>A manager who refers to policies, rules, procedures, obligation, or authority to influence the target individual</td>
<td>13</td>
<td>3</td>
<td>16</td>
<td>6</td>
</tr>
<tr>
<td>Reward</td>
<td>REW</td>
<td>A manager who offers instrumental rewards and favour to influence the target individual</td>
<td>12</td>
<td>4</td>
<td>16</td>
<td>6</td>
</tr>
<tr>
<td>Appraisal &amp; Control</td>
<td>A&amp;C</td>
<td>A manager who monitors and assesses the performance of usage by target individual</td>
<td>10</td>
<td>5</td>
<td>15</td>
<td>5</td>
</tr>
</tbody>
</table>
The item generation stage is probably the most important step in developing quality measures in terms of capturing the intended underlying concept. The combined theoretical and empirical item generation process should prevent important aspects of a construct from being omitted. It should assure that the representation of the ‘real’ phenomenon is captured. The rigorous sorting process and expert discussions should prevent redundancy and lay the basis for the factor structure. In the next step, the expectations of the properties of the scales are evaluated empirically.

Step 2: Scale Development

At this point, the set of 41 items is ready to be administered to a sample so that the psychometric properties can be assessed. Two aspects are distinguished: construct validity, which is an issue of operationalisation or measurement between constructs, and reliability which is an issue within a construct (Straub et al., 2004). Different quantitative heuristics or techniques are available to assess aspects of construct validity and reliability. For construct validity the focus lies on the fit of the measures with each other: the items should converge on a construct (convergent validity) but the constructs should diverge from each other (discriminant validity). Here, an exploratory factor analysis is performed on a data set. In order to assess reliability, several traditional methods exist to establish the accuracy and stability of the measures. Here, the internal consistency is calculated and other techniques are briefly discussed.

Study design

A survey instrument was developed and administered to purchasing professionals in order to collect data on the 41 items. The wording of the items was discussed with three experts, which led to some minor revisions of items. The questions were all of the form: “I experienced the following influence tactics”. The scales used for all items were 7-point Likert-type scales, ranging from “totally disagree” to “totally agree”.

The survey was administered to participants in an e-procurement seminar for practitioners. This sampling frame was selected because it is representative for the population for which the scale is developed. The survey asked respondents to report on their experiences with influence tactics in a situation where they have adopted a new technology. Respondents were requested to think of a situation in which they were an intended user of a new information technology that was introduced in their organization. The survey consisted of three parts. The 41 items were divided over the first two parts. In the first part, the respondents were asked to report on the perceived use of ‘soft’ influence tactics (PREC, MR, IP, COLL) in their organisations, followed by ‘harder’ tactics (LEGP, REW, A&C) in the second part of the questionnaire. The third part of the questionnaire contained questions on the background of the respondents. Both paper-based and on-line versions were administered. In the on-line version, items were presented in random order so as to prevent response pattern bias.

Since multivariate techniques are susceptible to sample size effects, larger sample sizes increase the confidence in assessing construct validity and reliability. Several item-to-sample ratio’s have been proposed. However, as long as there is sufficient intercorrelation, a smaller sample size can be used. A total of 54 responses were collected, just above the absolute minimum for factor analysis suggested by Hair et al. (1998). During the conference, five
assistants administered the survey using laptops. When all five laptops were being used by respondents, the paper-based version of the questionnaire was provided. There were no significant differences in responses between the paper based, English version (n=26) and on-line, Dutch version (n=28) of the questionnaire.

Scale construction

As mentioned before, exploratory factor analysis was used to identify latent constructs. Before this could be carried out, the data was analysed for missing data, outliers and the prerequisites for multivariate analysis. Four cases were omitted from data analysis because of high proportions of missing data, leaving a total of 50 observations. Overall missing data was limited (<10%) and no apparent distribution of missing data across cases (observations) could be identified. The distribution of missing data by variables (items) showed more missing variables in the second part of the survey, but no data imputations were deemed necessary. In addition, univariate, bivariate and multivariate outliers were analysed, leading to the omission of another observation. Finally, normality, homoscedasticity and linearity were assessed. These were slightly violated, which in itself is not harmful for factor analysis as long as sufficient intercorrelation is present in the sample.

Inspection of the data matrix showed sufficient correlations (> 0.30) between variables to justify the application of factor analysis. In addition, the Bartlett's Test of Sphericity –testing the null hypothesis that the correlation matrix is an identity matrix (all of the diagonal elements are 1 and all off diagonal elements are 0) – is rejected (p < 0.0005). Correlation among variables is therefore present. On the other hand, the degree of intercorrelation and the appropriateness of factor analysis using the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (MSA) are troubling. The MSA is a measure that varies between 0 and 1, where 1 describes a perfect prediction without error by the other variables. A value of 0.5 is a suggested minimum (Hair et al., 1998). Initial analysis shows a low MSA of 0.379.

Variables with a low individual MSA score and high partial correlations (based on the anti-image correlation matrix) are excluded from the analysis. Partial correlations of variables should be small as variables can be explained by the factors: if the partial correlation is high there is no ‘true’ underlying factor and factor analysis is inappropriate. Individual variables with an unacceptable MSA were excluded, resulting in a final set of 21 variables. Unfortunately, this process of item purification led to the deletion of all 5 items measuring Appraisal & Control. The overall MSA of the resulting set of variables is 0.643, which is well above the qualifying threshold of 0.5. An additional advantage of the reduced set of variables is the improvement in the sample-to-variable ratio, which contributes towards improved interpretation of findings in this relatively small sample.

A common factor analysis approach is selected a priori for the data analysis. There are two types of factor analysis: common factor analysis and (principal) component analysis. Both are popular data reduction techniques and if samples are large or communalities are high, both approaches show comparable results. Their overall goals, however, differ: component analysis primarily serves the purpose of data summarization, whereas common factor analysis focuses on understanding the structure. The purpose of common factor analysis is more ‘theoretical’ than ‘psychometric’. Component analysis uses common, specific and error variance, while common factor analysis only uses the common variance. The latter approach uses estimated
communalities of the variance of the variables. When prior knowledge suggests that specific and error variance is low, component analysis is appropriate (Hair et al., 1998). This knowledge is not available in this exploratory factor analysis. Moreover, common factor analysis estimates underlying constructs that cannot be measured directly and thereby provides a better basis for future use in SEM. Principal Axis Factoring is selected as extraction method. The extraction method is the algorithm to identify the loadings. Principal axis factoring is generally used for common factor analysis.

A visual assessment of the scree plot, assessment of eigenvalues and preliminary notions of the data structure all point towards the extraction of six factors. Cumulative variance extracted is 74%, which is above the suggested minimum of 60% for social science (Hair et al., 1998). Varimax rotation was selected as rotation method. Rotation helps simplify the factor pattern in order to achieve more meaningful theoretical factor structures by rotating the (reference) axis. Rotation also helps to redistribute the variance from earlier to later factors. Rotation techniques fall into two categories: orthogonal and oblique. With orthogonal rotation the extracted factors are uncorrelated by definition and in oblique rotation this property is relaxed and the factors are allowed to be correlated. The latter is more flexible and arguably closer to social reality. However, oblique rotations can become specific in a sample with a low case-to-variable ratio. For the purpose of identifying latent independent constructs correlation it is desirable that the independence of factors is preserved. Therefore an orthogonal rotation technique is chosen. There are several different orthogonal rotation techniques, e.g. varimax (attempts to achieve loadings of ones and zeros in the columns of the factor matrix ), quartimax (attempts to achieve loadings of ones and zeros in the rows of the component matrix), equimax (combines the objectives of both varimax and quartimax rotations). Varimax is chosen due to the clearer separation of factors. The rotated factor solution shows a simple structure with 21 variables loading upon 6 factors.

Interpretation of the factor matrix is based on an assessment of factor loadings. An important distinction for considering factor loadings is the difference between practical and statistical significance. Simply put, the larger the sample or the number of variables, the smaller the loading can be to be considered significant. Here, a small sample is used, so results should be interpreted with caution. Hair et al. (1998) show that in a sample of 50, a factor loading of 0.75 is required for significance of 0.05 (and power of 0.80). These are quite conservative guidelines and should be used as a starting point for interpretation (Hair et al., 1998). Hair et al. (1998) suggest guidelines for practical significance based on the notion that squared loadings are the amount of variance of the variable accounted for by the factor. Loadings of ±0.30 are considered to be the minimal level; loadings of ±0.40 are important and loadings of ±0.50 and larger are considered to be practically significant. The guidelines are based on a minimum sample of 100. Therefore, a more conservative threshold of ±0.40 was used here. All loadings above ±0.40 are shown in Table 2.

Reliability assessment
Reliability is a prerequisite for achieving valid measurements. The assessment should therefore be an integral part of scale construction and assessed iteratively for different factor structures. Reliability refers to the internal consistency of items in a construct and the stability of the items in construct over time. The most commonly accepted internal consistency measure is Cronbach's Alpha. Traditional stability measures include split-halves, test-retest and inter-rater
reliability. Here, only the internal consistency reliabilities are calculated. The generally accepted cut-off rate for Cronbach's Alphas in social science research is 0.70 (Nunnally, 1978). In exploratory research, however, Alphas as low as 0.50 have been deemed acceptable (Hinkin, 1998). One factor has a Cronbach Alpha of 0.688, while the other five range from 0.730 to 0.897. Several scholars tend to delete items from a construct to increase the coefficient Alpha, following the ‘purification’ guidelines in the Churchill approach. This practice, however, can jeopardize content validity, as will be discussed in step 3 below.

Table 2  Rotated factor matrix

<table>
<thead>
<tr>
<th></th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
<th>Factor 4</th>
<th>Factor 5</th>
<th>Factor 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>REW2</td>
<td>.799**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REW3</td>
<td>.744*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REW4</td>
<td>.711*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REW5</td>
<td>.831**</td>
<td>.833**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REW6</td>
<td>.745*</td>
<td>.797**</td>
<td>.783**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEGP2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.685*</td>
</tr>
<tr>
<td>LEGP3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.634*</td>
</tr>
<tr>
<td>LEGP5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.634*</td>
</tr>
<tr>
<td>LEGP6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.634*</td>
</tr>
<tr>
<td>IP2</td>
<td></td>
<td>.887**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IP4</td>
<td></td>
<td>.654*</td>
<td>.534*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IP5</td>
<td></td>
<td>.606*</td>
<td>.883**</td>
<td></td>
<td></td>
<td>.597*</td>
</tr>
<tr>
<td>COLL1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COLL2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COLL3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.634*</td>
</tr>
<tr>
<td>COLL4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.634*</td>
</tr>
<tr>
<td>PREC3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.405*</td>
<td></td>
</tr>
<tr>
<td>PREC5</td>
<td></td>
<td></td>
<td></td>
<td>.689*</td>
<td>.768**</td>
<td></td>
</tr>
<tr>
<td>PREC6</td>
<td></td>
<td></td>
<td></td>
<td>.839**</td>
<td>.678*</td>
<td>.627*</td>
</tr>
<tr>
<td>MR2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Extraction Method: Principal Axis Factoring. Rotation Method: Varimax with Kaiser Normalization. Rotation converged in 7 iterations. * Item is considered practically relevant, ** item is considered practically relevant and statistically significant (p<0.05)

Step 3: Scale Evaluation

The previous two steps were all about the development of a new scale. In step 3, the scale is tested using a new sample. A new sample is recommended so as to prevent sample specific findings of the factor structure. Naturally, iteration can take place between the evaluation and development steps. Straub et al. (2004) describe validation guidelines for scales and make a distinction between techniques for a) content validity, b) construct validity and c) reliability. Rossiter does not explicitly mention the necessity of an additional sample but stresses the a-priori specification and focus on content validity. The Churchill approach, on the other hand, explicitly recommends a second data set to assess reliability, i.e. the stability of the measure across
different measurements and preferably also different methods. In addition, construct validity should be reassessed. Churchill advocates the multi-trait multi-method technique to combine the aforementioned objectives. Confirmatory factor analysis techniques can also be used to test discriminant and convergence validity. Hinkin (1995) advocates other tests than purely ‘within-measure’ factor analysis and stresses the importance of demonstrating criterion-related validity (predictive or nomological validity). This can be done by confirming theoretically hypothesized relationships with regression analysis or structural equation modelling. Several additional validations techniques exist, like assessing the relationship between a scale and surrogate measures (criterion related validity).

Before the scale developed in this paper is ready for additional testing with a new sample, a careful reanalysis of content validity is deemed necessary. The data analysis resulted in the omission of nearly half of the items from the scale (20 out of 41). In the classic Churchill approach, the next phase would be gathering a new sample to test reliability and construct validity. In this case, however, a reassessment of content validity is added. The remaining scale could be damaged by the ‘cleansing’ of the scale on the basis of the psychometric properties. For each of the six factors, we evaluate whether the items that 'survived' the cleansing still adequately represent the original construct (i.e. influence tactic).

1. **Reward**
   The content validity of the items used to measure the construct 'reward' is satisfactory. The original conceptualization of reward is still fully reflected in the variables that load high on the first factor. Only one item was dropped. The factor analysis indicates that the following elements are part of the same underlying unidimensional construct: awarding (prize), rewarding and doing a favour. The construct of reward still has a predominantly instrumental nature; however, ‘doing a favour’ and ‘exchange’, which are still included, could also have a non-instrumental connotation of the scale.

2. **Legitimate pressure**
   The content validity of the items used to measure the construct 'legitimate pressure' is satisfactory. The original conceptualization is still largely reflected in the variables that load high on the second factor. Two items were dropped. Factor analysis indicates that the following elements of legitimate pressure are part of the same underlying unidimensional construct: using policies/rules/procedures and using authority. The fist seems to be a stronger part of the latent construct. Items that were dropped refer more explicitly to the use of authority and ‘pushing’ compliance.

3. **Information Push**
   The content validity of the items used to measure the construct 'information push' raises some concerns as it only partly covers the conceptual definition. The original conceptualization is only partly reflected in the variables that load high on the third factor. Three items were dropped that deal with the characteristics and performance of the system. Factor analysis indicates that the following elements of information push are part of the same underlying unidimensional construct: options of the system and ways to operate the system. The first could implicitly refer to information to form a notion of the system and its performance, however, the three item construct shows a stronger emphasis on information to support learning how to operate the system. A theoretical basis exists
for factual information supporting both the assessment of the utilitarian aspects as well as the usage/learning aspects (e.g., two types of rational persuasion as defined by Yukl). Nomological validity can be jeopardized with a minimal element of performance related information. As a remedy, it is suggested to add an item focusing on the system characteristics (IP1) in order to balance the content of the construct.

4. **Collaboration**
The content validity of the items used to measure the construct collaboration is satisfactory, although it is more specified: the source is limited to an expert and the application domain is limited to the working context and/or purchasing situation. The original conceptualization is partly reflected in the variables that load high on the fourth factor. Three items were dropped that deal with identifying opportunities, expert help and a manager searching for input. Seeking an opinion could be seen as part of a different construct; the other items only repeat and generalize the items that are retained. Factor analysis indicates that the following elements of collaboration are part of the same underlying unidimensional construct: jointly determining the suitability, specification, and use. These elements reflect three consecutive phases. The emphasis on the purchasing context is not a problem as generalizing outside the purchasing domain is not an explicit goal for this research. Reflection on these findings leads to the redefinition of the concept as collaboration with expert, instead of augmenting the scale to include a managerial role.

5. **Persuasive Recommendation**
The content validity of the items used to measure the construct persuasive recommendation is somewhat troubling. The original conceptualization is only partly reflected in the variables that load high on the fifth factor. The element of persuasion by converting opinions and countering hesitations was dropped as a result of the factor analysis. Three items were dropped that deal with converting, persuasion with arguments, benefits for the ‘big picture’ and countering hesitations. The role of information (argumentation) was already a confusing issue for raters in the clustering effort (related to the construct information push): this led to adding ‘factual’ in IP and leaving the possibility for normative information in PREC. After factor analysis the emphasis on inspirational persuasion, i.e., normative information, is even stronger. Factor analysis indicates that the following elements of persuasive recommendation are still part of the same underlying unidimensional construct: recommendation for business benefit and inspiring. The emphasis (in terms of loadings) is on the latter. The implication of the factor loadings is that the actual recommendation part for personal benefit is diminished. To mend this issue the conceptual definition is altered.

The reliability of the 3 item-scale for persuasive recommendation is below the recommended 0.70 (Cronbach $\alpha$ of 0.688). Omission of item PREC3, however, would increase the reliability of the scale to a satisfactory level (Cronbach $\alpha$ of 0.779). This would further increase the shift of the underlying concept from recommendation to a part of persuasion. The construct thereby comes close to the construct inspirational appeal (Yukl, 2001). The initial conceptualization is better preserved with a three item construct and therefore preferred.
6. Managerial Request
The content validity of the items used to measure the construct managerial request is satisfactory. The original conceptualization is fully reflected in the variables that load high on the sixth factor. Three elements were dropped that deal with asking, requesting cooperation, and stating wishes. Although the latter could be seen as distinctive aspect, the initial conceptualization of a request remains intact. Factor analysis indicates that the following elements of managerial request are still part of the same underlying unidimensional construct: requesting compliance, requesting acceptance of ideas and asking politely. The emphasis (in terms of loadings) is on the first element.

7. Appraisal & Control
The A&C construct was not confirmed in the exploratory factor analysis. This implies that the items initially expected to form the A&C construct did not load on one dimension in the data sample. The conceptual definition in relation toward the other influence tactics should therefore be seriously revised to be considered for inclusion in a next round of data collection. Based on the previous inter-rater reliability in step 1, it is suggested to limit the construct towards control of a target’s behaviour. Only two items (A&C1 and A&C3) are then included.

Balancing reliability and content validity
Using empirical data to assess the quality of scales presents some caveats. A clean structure in the factor loading does not guarantee that the content validity, as determined ex ante, is still intact. It is important not to take psychometric findings and factor structure as they are, but to review if the factors still represent the intended concept. “Simply because items load on the same factor does not imply that they necessarily measure the same theoretical construct” (Nunnally, 1978). A comparable logic also holds for the assessment of reliability: an acceptable internal consistency does not guarantee that the ‘right’ concept is measured, only that it is measured accurately. The discussion of the content analysis showed that four constructs remain intact (REW, LEGP, COLL, MR), that the PREC construct is slightly redefined and additional items are proposed for IP. The A&C is completely redefined, because no factor emerged from exploratory factor analysis.

In total, a revised list of 24 items, representing seven underlying dimensions, is now ready to be submitted to a second, preferably larger, data set to be tested. The analysis should include a confirmatory factor analysis (CFA) and assessment of criterion-related validity by assessing the relationship of the constructs with (previously used) dependent variables.

This section discussed a balancing act between the output of the first round of data collection on the developed scale and the initially determined conceptualizations of the underlying constructs. The discussion of content validity leads to a set of 24 items which is ready for another round of data collection. This second round of data collection is not discussed in this paper (data collection is currently being performed). In the next section, the implications of the proposed and demonstrated method for scale development are discussed.

Discussion
In this paper, a three-step approach is proposed for scale development, balancing Churchill’s and Rossiter’s classical view. Rossiter’s C-OAR-SE procedure relies primarily on content
validity, whereas Churchill focuses on construct validity and its assessment of psychometric properties. Here, the focus on content and construct validity is balanced. The three steps are: (1) item generation, (2) scale development, scale construction and reliability assessment, and (3) scale evaluation.

This paper provides an application of the proposed method for the development of a scale for seven influence tactics. First of all, items are generated from both existing research and interviews on influence tactics using a theoretical framework as a guiding framework. The items are submitted to a rigorous sorting process using 19 raters, condensing the long list of 104 to 41 items. A first data set (N=54) was collected during an EP seminar and an exploratory factor analysis was performed. This resulted in a set of 21 items representing 6 underlying factors. Before the scale could be evaluated with a second data set, content validity has been reassessed. This resulted in a proposed 24-item scale for 7 influence tactics, that still needs to be tested.

This paper initiates the discussion for better scale development in purchasing research and enhances our understanding of influence tactics in EP adoption. A quick scan of purchasing literature, proves that limited interest has been demonstrated for the development of quality scales. The processes for scale development applied in purchasing research are generally limited to using “face-valid” measures or a mix-and-match of previously used scales without additional testing to see if they still adhere to validity and reliability criteria. In addition, evaluating previously used scales is generally limited to psychometric tests. This emphasis on statistical analysis may underestimate the importance of accuracy of measurement (Hinkin, 1995). A purely inductive approach, on the other hand, may overemphasize the contextual nature and put limits on the generalization of the scale. The approach presented here shows a way in which development and evaluation can strike a balance between these approaches and to come closer towards capturing the ‘true’ concept of interest.

The suggested approach and application does not provide a definitive framework or approach, but resonates the call, that has been made previously in other disciplines, for more attention to developing high-quality scales in the field of purchasing. This paper merely takes a first step in this direction.

References