

Validation of Scales of Measurement in Administration

Validação de Escalas de Mensuração em Administração

Eric David Cohen*

Centro Universitário Alves Faria, Goiânia, GO, Brazil

ABSTRACT

This article presents a discussion related to the empirical techniques used in the construction of theories in Business Administration, focusing on the confirmatory methods. An important theoretical contribution is the proposed categorization of the research, specifically: incipient knowledge, knowledge with conceptual definitions, and consolidated research knowledge, through the validation of the measurement scales. In this research, two classic scales of Organizational behavior were used to illustrate the discussion of "theory destruction" (Bido, 2014). Specifically, the replication of the scales ought to be carried out with a set of empirical procedures that verify the validity of the conceptual model, which takes into account the data collected. From a common methodological perspective, the issues pertaining to the construction of theory emerge in light of the fact that the original conceptual structure has not been preserved. In addition, there was unwarranted use of the exploratory technique (instead of the confirmatory technique). Confirmatory methods assist researchers and reviewers in following proper research practices, which could lead to the advances in the theory, or alternatively point out modifications in the constructs, based on the adjustment indices of the measurement models.

KEYWORDS: Measurement scales; Validation of scales; Confirmatory factor analysis; Theory construction; Quantitative methods.

RESUMO

Este artigo analisa as técnicas empíricas utilizadas na construção de teorias da área de Administração, com foco nas técnicas confirmatórias. Apresenta-se como contribuição teórica importante, um esquema de categorização da pesquisa a partir do: conhecimento incipiente, conhecimento com definições conceituais e conhecimento acumulado das pesquisas pela validação de escalas de mensuração. Para a replicação de escalas, é necessário empregar procedimentos empíricos que verifiquem a validade do modelo conceitual, frente aos dados coletados. Na presente pesquisa, foram analisadas duas escalas clássicas da área de Comportamento Organizacional, as quais foram replicadas por autores nacionais, com o fito de ilustrar a discussão da "destruição de teoria" (Bido, 2014). A partir de uma perspectiva metodológica comum, emergem questionamentos em relação às escalas analisadas, posto que, em ambos os casos, não se apresentou uma justificativa de mudança em relação à estrutura conceitual original, sendo utilizada - de forma inadequada - a técnica exploratória (em vez da técnica confirmatória). Esta pesquisa corrobora a necessidade de aderência aos métodos confirmatórios, tanto para os pesquisadores quanto para os revisores de artigos científicos, para justificar possíveis revisões das teorias ou para indicar modificações nos construtos por meio de índices de ajuste dos modelos.

PALAVRAS-CHAVE: Escalas de mensuração; Validação de escalas; Análise Fatorial Confirmatória; Construção de teoria; Métodos quantitativos.

Submission: December 22, 2017

Approval: April 18, 2018

*Eric David Cohen

Post-Doctorate in Administration (Quantitative Methods) by Universidade Mackenzie. Professor of the master in business Administration from Centro Universitário Alfa, and Professor of Faculdade de Ciências Aplicadas da Universidade Estadual de Campinas (Unicamp), in the areas of Market Research, Strategy and Quantitative Methods.

Address: Av. Perimetral Norte, n. 4129, 74445-190, Goiânia, GO, Brazil.

E-mail: EricDCohen@gmail.com

1 INTRODUCTION

The application of analytical techniques and measurement scales has made a significant contribution to the construction of theory in Business Administration. For instance, in Organizational Behavior, we can mention the measurement scales for organizational climate, learning and commitment to the company, among numerous other themes in the field.

Despite its importance in theory construction, there are gaps in the use of suitable analytical methods and algorithms. In that respect, researchers such as Conway and Huffcut (2003), Fabrigar, Wegener, Maccallum and Strahan (1999) and Hinkin (1995) posit that – sometimes - researchers use analytical techniques in an improper fashion, which contradicts the extant theory.

This issue was specifically addressed in the Brazilian Forum for Teaching and Research for the Business Administration and Accounting area, under the auspices of the National Council of Graduate Program in Business Administration (ANPAD). During this event, there were abundant examples of research that ineffectively used analytical techniques. Confronted with this problem, researchers reached consensus with regards to the effective use of the factor analysis techniques (Bido, 2014).

Along the same vein, Bido, Ribeiro and Cohen (2016) conducted a bibliometric study of national and international Business Administration journals of higher strata, ranging from 2010 to 2015. The sample of national surveys revealed an inadequate use of analytical methods and inconsistent interpretation of results, which confirmed the ineffectual configuration of scale construction methods and theory development in the field. Given the previous knowledge of the quantity and meaning of the constructs, the unwarranted use of exploratory methods leads to diverging factor structures. In such cases, the recommended course of action is to use confirmatory factor analysis or structural equation modeling, in lieu of the exploratory technique.

Hence, the purpose of this study is to analyze research practices in the validation of measurement scales, and to discuss the transference of scales created in different cultures and languages – a complex task that sometimes results in contradictions.

It is worth mentioning academia's efforts to discuss analytical techniques and best practices. These are evidenced by the community's efforts to improve research methods, such as the Academy of Management and the Academy of International Business divisions for the discussion of research methods, and periodicals such as the Organizational Research Methods journal.

2 RESEARCH DEVELOPMENT

Hinkin's (1995, 1998) contribution to theory construction was grounded in Hunt's inductive and deductive approaches (1991). According to Hinkin, the inductive approach occurs in an exploratory manner, whereby the researcher attempts to develop a conceptual model that emanates from the data that was collected. In contrast, the deductive approach begins with a theoretical framework and its conceptual definitions that guides the research, along the generation of items from an existing measurement scale.

In the following section, we present a review of the analytical techniques and recommended practices for the construction of measurement scales.

3 LITERATURE REVIEW

3.1 Factor Analysis

Factor Analysis is a statistical technique which evaluates the variability of manifest variables (henceforth, items) that compose a smaller number of latent variables that cannot be measured directly. This is considered a data reduction method, since it identifies the redundancy underlying a set of items (Hair, Black, Babin, Anderson, & Tatham, 2009). The technique is subdivided into two types: Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA).

The former uses an inductive approach to investigate the relation between the factors and the items; it is generally used in the initial phases of theory construction. Here, there are no prior assumptions about the conceptual structure (or dimensionality) of the constructs (Jöreskog, 2007; Netemeyer, Bearden, & Sharma, 2003; Pett, Lackey, & Sullivan, 2003). Along the process, the researcher starts with the collection of data and seeks a solution that best represents some conceptual structure. The EFA identifies the items that do not contribute for the measurement of the construct, or those that present cross-factorial loadings (Fabrigar et al., 1999, Pasquali, 2012; Worthington & Whittaker, 2006).

In contrast, the Confirmatory Factor Analysis (CFA) is characterized by a deductive approach, based on the theoretical knowledge that underlies some specific phenomenon. This approach presupposes that the existing scales have gone through the stages of scale construction and refinement, which culminates in the construction of theory. It is thus assumed that there are prior conceptual specifications regarding the number of factors, as well as the indicators that make up each factor (Peters et al., 2003; Worthington & Whitaker, 2006). The researcher must make use not only of the scale itself, but also of the theory that supports it, as well as extant research.

The confirmatory techniques offer a set of adjustment indices that summarize the *post-hoc* results of empirical model tests. The most widely used metrics are the χ^2 , the normed χ^2 and construct, convergent and discriminant validity (Hair et al., 2009; Hinkin, Tracey, & Enz, 1997) (see attachment).

3.2 Theory destruction

The term "destruction of theory" was coined by Bido (2014) to denote the situation whereby the results differ from previous research, because of improper use of exploratory methods in existing scales. This presents an issue, because the theory that underlies the research has not been preserved. Since EFA extracts factors without considering any conceptual structure, it ends up generating groups of items in a random manner.

In other words, the application of the EFA to validate (or replicate) a scale may lead to a questionable research decision, by introducing modifications in the original scale. It could be argued that the theory has been destroyed, since it has been left aside. In addition, the possibility of comparing results with previous research is lost, not to mention that the theoretical sense of the constructs may be jeopardized.

In view of the research objectives, it is important to note that the replication of a validated scale is not always warranted. Some points to be considered are: possible interpretation problems by the target audience; choice of scale development methods; and cultural and context differences. In all cases, it is important to emphasize the need to calculate the adjustment metrics, and to ensure that the scale is valid and reliable.

3.3 Recommended procedures for the construction and refinement of scales

In the process of constructing theory, the researcher may elect to make significant modifications to the scale; alternatively, s/he may consider developing a new scale. At any rate, there are recommended procedures for building, developing and refining scales.

Authors such as Costa (2011), Devellis (2016) and Netemeyer, Bearden and Sharma (2003) suggest specific criteria during the stages of theory construction. These authors postulate different courses of action based on previous knowledge and theoretical reference (Figure 1).

Hinkin (1998)	Devellis (2016)	Costa (2011)	Netemeyer et al. (2003)
<p>Step 1: Generation of items: (i) deductive approach; (ii) inductive approach; (iii) development of items; (iv) evaluation of content validity; (v) number of items; (vi) scale type (agreement, frequency, etc.)</p> <p>Step 2: Administration of the questionnaire: sample size</p> <p>Step 3: Reduction of items: Exploratory Factor Analysis and reliability</p> <p>Step 4: Confirmatory Factor Analysis</p> <p>Step 5: Convergent and discriminant validity</p> <p>Step 6: Replication</p>	<p>Step 1: Clearly determine what is intended from the measurement scale: (i) theory; (ii) specificity; (iii) which items to include</p> <p>Step 2: Generate items pool: (i) choose items that reflect the purpose of the scale; (ii) redundancy; (iii) number of items; (iv) writing; (v) bad and good items; (vi) reverse items</p> <p>Step 3: Determine the measurement format: (i) Thurstone; (ii) Guttman; (iii) equally weighted items; (iv) number of response categories; (v) type of response format (agreement, semantic differential, etc.)</p> <p>Step 4: Initial pool of items reviewed by experts</p> <p>Step 5: Include items for validation (social desirability, criterion validity)</p> <p>Step 6: Administer the items for a pre-test sample</p> <p>Step 7: Evaluate items</p> <p>Step 8: Optimize scale size (use item-total correlation and reliability to evaluate items; EFA is recommended first to ensure that items are one-dimensional)</p>	<p>Step 1: Construct domain specification.</p> <p>Step 2: Item generation, face and content validation.</p> <p>Step 3: Decisions about the responses</p> <p>Step 4: Construction of the research instrument</p> <p>Step 5: First sampling</p> <p>Step 6: first round scale cleaning, (EFA, correlation, reliability)</p> <p>Step 7: Additional fieldwork</p> <p>Step 8: Scale cleanup round (same as step 6 + CFA)</p> <p>Step 9: Validity and reliability of the final scale</p> <p>Step 10: Develop standards and recommendations (how to apply scale and interpret scores)</p>	<p>Step 1: Definition of the construct and content domain: (i) reflexive items and formative indicators; (ii) role of theory in specification of dimensionality</p> <p>Step 2: Generation and decision of measurement items: (i) domain sampling; (ii) generation of the items pool; content and face validity</p> <p>Step 3: Plan and implement work to develop and refine the scale: (i) pilot test; (ii) initial EFA; (iii) reliability</p> <p>Step 4: Finalize scale development: (i) EFA; (ii) CFA; (iii) predictive and concurrent validity; (iv) establish rules (to interpret the scores and show that the means of the scores are different between groups that are admittedly different)</p>

Figure 1 – Guidelines for the construction of measurement scales

Source: Adapted from Hinkin, 1998, Devellis, 2016, Costa, 2011, Netemeyer et al., 2003.

Figure 2 summarizes the references from the perspective of the degree of knowledge. It may therefore be used to guide the discussion of measurement scale construction and validation. It should be noted that these references are not clearly explained by the authors; therefore, these guidelines are an important contribution of this paper.

In the first scenario, there exists an incipient knowledge which usually translates into an exploratory or qualitative research approach. In the second case, there is accumulated knowledge arising from the conceptual definitions and the construct dimensionality; however, there is some uncertainty regarding the concept structure and validity. The last scenario depicts the existence of a broad body of accumulated scientific knowledge.

Authors like Conway and Huffcut (2003) and Fabrigar, Wegener, Maccallum and Strahan (1999) provide very relevant definitions with regards to the best practices in EFA. Yet, they fail to provide a discussion regarding the use of exploratory techniques.

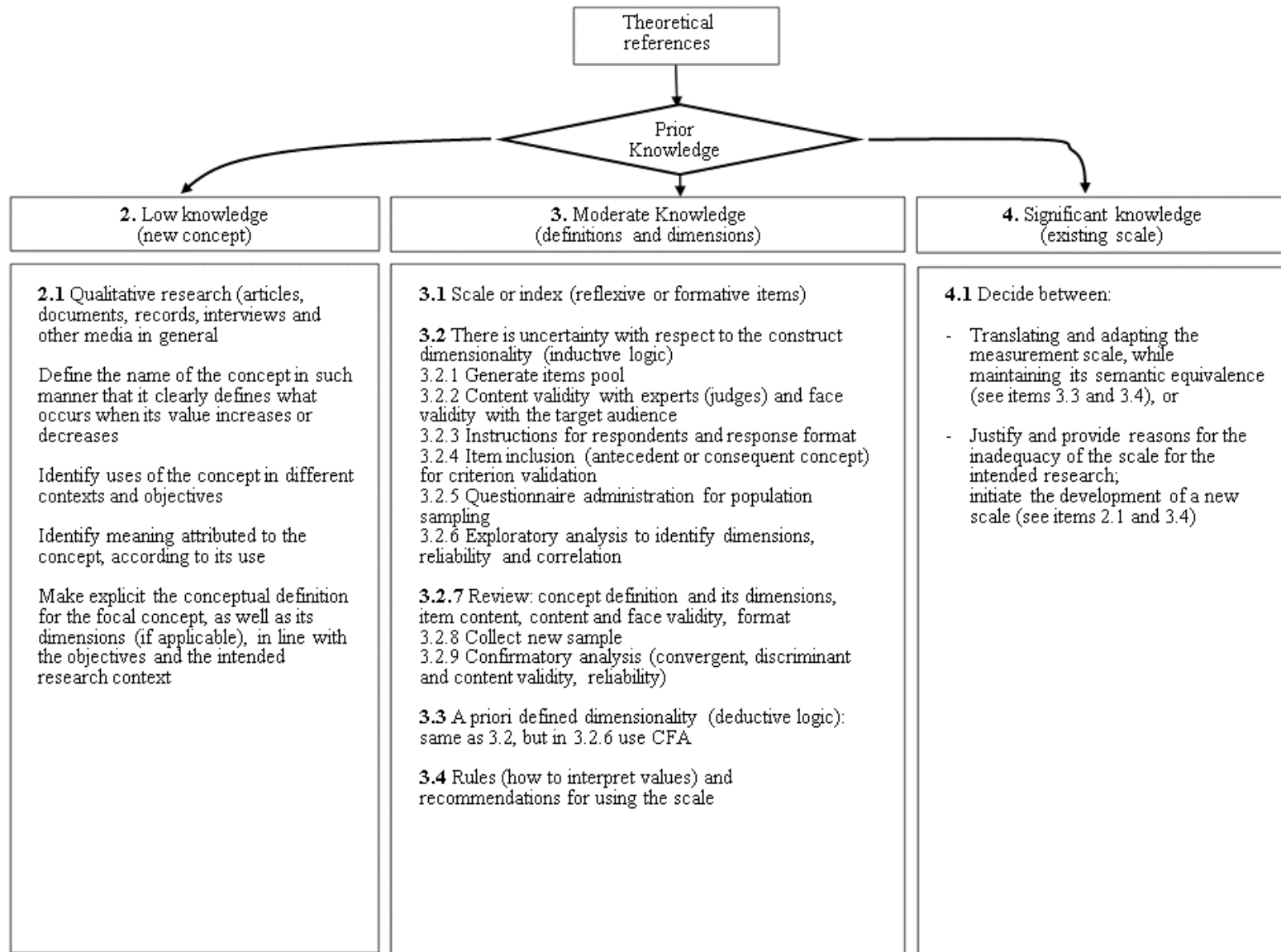


Figure 2 - References related to the degree of knowledge
Source: Prepared by the author.

The use of sound methodological practices – which respect the theoretical framework - help circumvent questions related to the application and interpretation of theory. Best practices preserve the theory's academic and managerial relevance, and greatly benefit the scientific publications referees and researchers.

The relevance of this research with respect to Hunt's (1991) criterion thus becomes more evident: after discussing these issues, what type of knowledge with regards to the construction of measurement scales will be generally available to the academic community in the field, which is not known today?

4 RESEARCH METHODOLOGY

To support the discussion of the central research problem, two measurement scales based on the classic theories of Organizational Behavior were selected. These scales were published by Siqueira (2008) and replicated by Brazilian researchers – however, the original conceptual structure was not preserved (Table 1).

The choice of these constructs for our research purposes is predicated in the fact that there are conceptual differences between the original and the replicated scale. In other words, the criterion for consideration is based on the possibility that the chosen scales present different conceptual structures, which result from using improperly an exploratory technique¹. The methodological approach adopted in this paper uses confirmatory methods to analyze the underlying conceptual structure. Its results may either suggest the need for a review of the theory, or corroborate previous research findings.

Table 1 – Organizational behavior scales

Construct	Brazilian scale			Original scale		
	Authors	Factors	Items	Authors	Factors	Items
Involvement with work	Siqueira, 2008	1	5	Lodahl and Kejner, 1965	3	20
Work values	Pasquali and Alves, 2004	4	40	Schwartz, 1992	10	40

Source: Prepared by the author.

Both measurement scales were subjected to the stages of translation and validation, verification of sampling and data collection requirements and application of confirmatory techniques. The analysis of results is presented along with validity and reliability metrics.

4.1 Sampling procedures

4.1.1 First study: Involvement with work

Lodahl and Kejner's (1965) questionnaire for Involvement with work was implemented in Google Forms. Data collection took place in the second semester of 2016, yielding 231 effective responses from the employees of a multinational company that industrializes white goods; the company is based in the countryside of the State of São Paulo.

The scale factors are presented in Figure 4. The items correspond to Likert-type questions in a five-point scale of agreement, ranging from 1 (totally disagree) to 5 (completely agree).

The confidentiality of the responses was assured, and the employees received an email invitation from management representatives. The respondents were from various functional areas, gender and hierarchical levels.

Inspection of the sample response values reveals that there was no missing data; therefore, the use of imputation methods to treat missing values was not required. The sample size complies with the recommendations by Hair, Black, Babin, Anderson, & Tatham (2009), as well as Rosseel's (2012)

¹ This research is currently being expanded to incorporate other Organizational behavior constructs, with the intent of publishing a book that presents the best practices in scale development.

suggestion for using the "WLSMV" estimator – which is default for categorical variables; this estimator uses standard errors based on polycyclic correlations of bivariate data.

4.1.2 Second study: Work values

Based on the contributions of Pasquali and Alves (2004), a questionnaire was constructed using Google Forms, with information related to the work values of employees. Data was collected in the second semester of 2016 from employees of a multinational company, located in the countryside of the State of São Paulo. The respondents were ensured of data confidentiality, and received an email invitation from management for participation in the survey (N = 200).

In the questionnaire, respondents were asked to indicate the degree that a specific situation resembles a hypothetical person. The scale factors are presented in Figure 3; items were collected using a six-point scale, ranging from 1 (does not look at all like me) to 6 (looks a lot like me). Respondents came from several different functional areas, ages, gender and hierarchical levels. No missing data was found in the sample.

Question	Item
1	Thinking about new ideas and being creative is important to him/her. S/he likes to do things in his/her own original way.
2	Being rich is important to him/her. S/he wants lots of money and own expensive things.
3	S/he believes it is important that all people in the world are treated equally. S/he believes that everyone should have equal opportunities in life.
4	It is very important for him/her to demonstrate his/her abilities. S/he wants people to admire what s/he does.
5	It is important for him/her to live in a safe environment. It avoids anything that could put your safety in danger.
6	S/he thinks it's important to do several different things in life. S/he always looks for new things to try.
7	S/he believes that people should do what they are told to do. S/he believes that people should always follow the rules, even when no one is watching.
8	It is important for him/her to listen to people who are different from him/her. Even when you do not agree with them, you still want to understand them.
9	S/he thinks it's important not to want more than you have. S/he believes that people should be satisfied with what they have.
10	S/he looks for every opportunity to have fun. It is important for him/her to do things that give his/her pleasure.
11	It is important for him/her to make his/her own decisions about what s/he does. S/he likes to be free to plan and choose his/her activities.
12	It is very important for him/her to help the people around him/her. S/he wants to take care of their welfare.
13	Being very successful is important to him/her. S/he likes to impress other people.
14	The security of your country is very important to him/her. S/he thinks the government must be alert to threats of internal or external origin.
15	S/he likes to take a chance. S/he is always looking for adventures.
16	It is important for him/her to always behave correctly. S/he wants to avoid doing anything that people might find wrong.
17	It is important for him/her to be in charge and tell others what to do. S/he wants people to do what s/he says.
18	It is important for him/her to be faithful to his/her friends. S/he wants to focus on people close to him/her.
19	S/he firmly believes that people should preserve nature. Caring for the environment is important to him/her.
20	Being religious is important to him/her. S/he strives to follow his/her religious beliefs.
21	It is important to him/her that things are organized and clean. S/he really does not like things to be messy.
22	S/he thinks it's important to show interest in things. S/he likes to be curious and try to understand all sorts of things.
23	S/he believes that everyone in the world should live in harmony. Promoting peace among all groups in the world is important to him/her.

24	S/he thinks it's important to be ambitious. S/he wants to demonstrate how capable s/he is.
25	S/he thinks it's better to do things in a traditional way. It is important for him/her to keep the customs s/he has learned.
26	Taking advantage of the pleasures of life is important to him/her. S/he likes to pamper him/herself.
27	It is important for him/her to understand the needs of others. S/he tries to support those s/he knows.
28	S/he believes s/he should always respect his/her parents and elders. It is important for him/her to be obedient.
29	S/he wants everyone to be treated fairly, even those s/he does not know. It is important for him/her to protect the weakest in society.
30	S/he likes surprises. It is important for him/her to have an exciting life.
31	S/he tries not to get sick. Being healthy is very important to him/her.
32	Progress in life is important to him/her. S/he strives to do better than others.
33	Forgiving the people who hurt him/her is important to him/her. S/he tries to see what is good in them and not to have a grudge.
34	It is important for him/her to be independent. S/he likes to count on him/herself.
35	Having a stable government is important to him/her. It is concerned with the preservation of social order.
36	It is important for him/her to be always polite to others. S/he tries never to bother or irritate others.
37	S/he really wants to enjoy life. Having fun is very important to him/her.
38	It is important for him/her to be humble and modest. S/he tries not to draw attention to him/herself.
39	S/he always wants to be the one to make decisions. S/he likes to lead.
40	It is important for him/her to adapt to nature and fit into it. S/he believes that people should not change nature

Figure 3 - Work values scale items

Source: Prepared by the author, adapted from Pasquali and Alves, 2004.

4.2 Analytical techniques

CFA and Structural Equation Modeling were used to support our research objectives – namely, to empirically test the measurement scales. According to Hair, Gabriel and Patel (2014), confirmatory techniques that use covariance data are suitable for testing theories, because they enable us to compare the scale in a different context and to identify its compatibility. If the scale measures the theoretical construct in a similar manner when replicated in a different context, it can be assumed that there is measurement scale invariance (Borsa, Damásio, & Bandeira, 2012).

The sample size requirements in both studies were impaired, because the measurement scales contain categorical response items. This means that possibly there is a limited set of values, data asymmetry and absence of values in certain points of the scale. As a result, the variability of the items is reduced; this, in turn, affects the statistical technique's performance and the convergence of the model (since the technique presupposes multivariate normality). To address that issue, we used the R software with Lavaan package (R Core Team, 2016).

5 DISCUSSION OF THE RESULTS

5.1 First study: Work involvement scale

The concept of Work involvement incorporates sociological aspects, as well as conditions that are present in the organization. Seen in this way, Work involvement contemplates social norms and values that affect the specific forms of behavior, which derive from the internalization of the perceived values with respect to the quality of the work, as well as the importance of the work and individual's self-esteem (Siqueira, 2008).

Throughout this study, Lodahl and Kejner's (1965) model provided the theoretical reference for our research. Initially, the authors identified 110 items, relying on the accumulated knowledge of a number of experts in the field. After eliminating redundant items, they reduced the scale to 87 items, which were subsequently presented to a group of experts (composed of eleven Psychologists, three Sociologists, and eight Human Relations scholars). They recommended the elimination of 47 items,

leading to the proposition of a scale with 40 items of four-point Likert type. This was applied to a group of 137 Nurses and 70 Engineers, and after successive refinements, the final scale was comprised of 20 items was proposed.

Lodahl and Kejner (1965) used an exploratory technique and found three underlying factors which were named Involvement Intensity, Work Indifference and Feeling of Pride and Professional Responsibility (Figure 4).

In the Brazilian context, Siqueira (2008, 2014) replicated Lodahl and Kejner's (1965) scale. Initially, it was comprised of eight items that use a seven-point Likert scale. After successive refinements, the author proposed a one-dimensional scale composed of five items: "the greatest satisfaction in my life comes from work"; "The most important things that happen in my life involve my work"; "I am personally very attached to my work"; "I eat, live and breathe my work"; and "the hours I spend working are the best hours of the day".

It should be noted that the last two items did not belong to the original Lodahl and Kejner (1965) scale. Unfortunately, Siqueira did not disclose the reasons that supported the inclusion of these two items; nor did she provide information pertaining to the construction of the measurement scale. As a result, the reader does not know whether the scale construction decisions derive from some issues pertaining to the item contents, or whether they are related to the theoretical framework. It is also relevant to note that the author does not cite the critical reviews of the original scale, from authors such as Saleh and Hosek (1976) and Reeve and Smith (2001).

The original scale depicts a three-dimensional construct, which is manifest through 20 items (Lodahl & Kejner, 1965). In contrast, Siqueira (2008) proposes a one-dimensional scale with five items. From a theoretical point of view, an important question regarding the construct dimensionality arises: is it one-dimensional, or multidimensional?

We need to verify whether the distinctions between the original and the replicated scale are related to conceptual differences that arise from cultural differences, or whether they result from the improper use of the exploratory technique. To that end, we will investigate whether the "destruction of theory" has occurred.

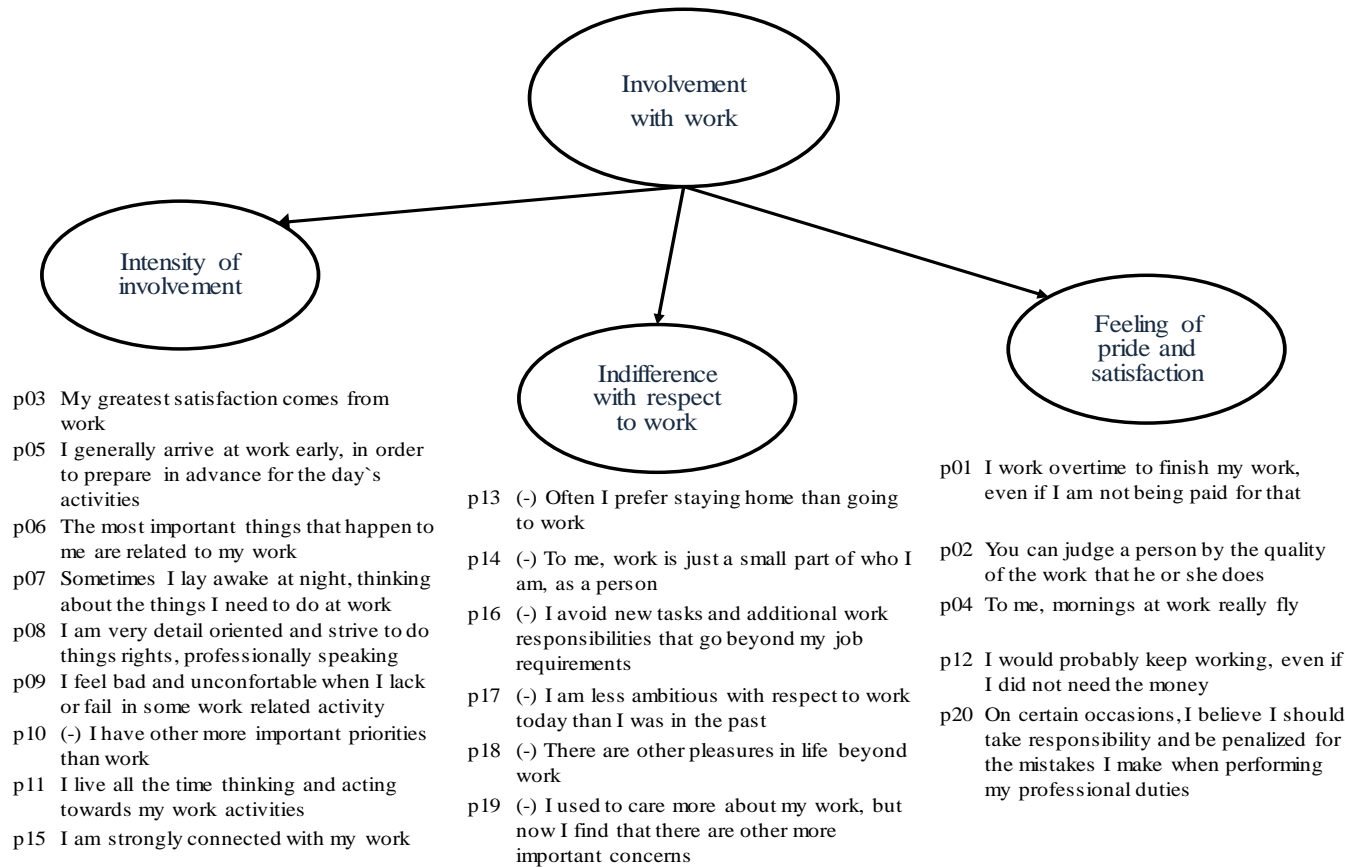


Figure 4 - Lodahl e Kejner (1965)'s scale for measuring Work Involvement
 Source: Prepared by the author, based in Lodahl and Kejner's (1965) work.

5.1.1 Empirical tests and scale refinement of the conceptual model

Confirmatory factor analysis was used to empirically test Lodahl and Kejner's (1965) scale. The factors were tested individually; however, this yielded unsatisfactory adjustment indices. The following items presented problems: in the first, p07 is not significant, and p15 has a high modification index. In the second, p16 and p18 are not significant. No problems were found in the third factor.

Hence, the original scale could not be confirmed. This result is in line with the issues raised by Reeve and Smith (2001) regarding the original scale's performance. An alternate model was tested:

$$F1 = \{p01+p02+p04+p08+p09+p15+p16+p17\}$$

$$F2 = \{p03+p06+p10+p14+p18+p19\}$$

$$F3 = \{p05+p07+p11+p20\}$$

The third factor was discarded due to non-significant loadings for items p07 and p20, as well as a high correlation between it and the second factor. The new model presents excellent adjustment indices: CFI = 1; RMSEA = 0; CI = 0.021; SRMR = 0.052; GFI = 0.998, normed $\chi^2 = 17.6$ and significant factor loadings. The residual matrix data is well within the recommended levels.

The convergent validity of the model in Figure 5 was analyzed, yielding an extracted mean variance for the first factor of 0.5090 and composite reliability of 0.8761. Both are within the parameters recommended by Hair et al. (2009) (≥ 0.5 and ≥ 0.7 , respectively).

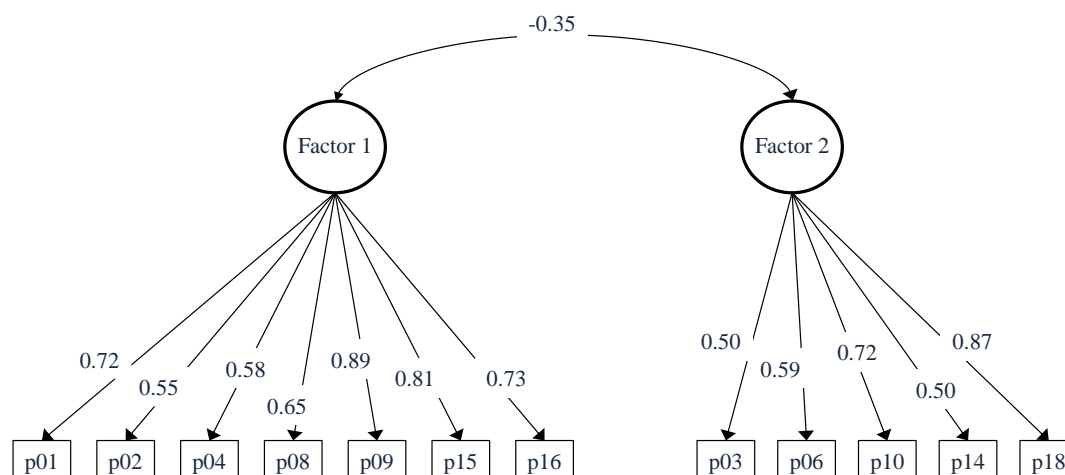


Figure 5 –Measurement scale for the refined Work involvement model

Source: Prepared by the author.

In turn, the second factor's extracted mean variance was 0.4244 and its composite reliability was 0.7783. Removing items p03 and p14 would increase convergent validity (taking the extracted mean variance to 0.5413, within recommended levels). Even though the elimination of items with low factorial loadings increases the extracted mean variance, this would jeopardize content validity and reduce the capacity to generalize the results. Moreover, in this situation we would be switching from a deductive approach (i.e. start from theory and empirically test using confirmatory methods), to an inductive approach (i.e. we start from the data collected and try to construct theory using an exploratory technique), and the decision was made to preserve the original model.

For the analysis of discriminant validity, we first calculated the value of χ^2 for the free model, and subsequently repeated this for a second model, where the correlation between the factors F1 and F2 is fixed to one. The calculated χ^2 are 41.67 and 227.84, respectively. We therefore reject the null

hypothesis that the correlation between the factors has a unit value and conclude that the model has discriminant validity.

A consequent construct was included to evaluate the construct criterion validity. At this stage, we answer the demand for utility: (a) what is the scale used for? (b) is the endogenous variable related to other variables? (Armstrong & Soelberg, 1967, Netemeyer et al., 2003, Osborne, 2014).

To that end, we looked for new conceptual definitions pertaining to the Work involvement concept, which do not comprise Lodahl and Kejner's (1965) construct definition. Brown (1996) presents a set of antecedent, correlated and consequential variables for Work involvement. Based on his contribution, the decision was made to include the variable "relationship with the manager" as a dependent construct for Work involvement:

p21 - The greater the Work involvement, the more intense will be the relationship with management.

The variable's criterion validity analysis – which considers the dependence from the exogenous variables F1 and F2 - is depicted in Table 2.

Table 2 - Results from the structural model

	Unstandardized structural coefficient	p(>z)	Standardized structural coefficient	R ²
F1->F3	0.368	0.000	0.347	11.1%
F2->F3	0.203	0.027	0.191	

Source: Prepared by the author.

The adjustment indices of the structural model are similar, leading us to conclude that an increase in the importance of work and professional pride will yield a moderate increase in the employee's relationship with management. This result confers criterion validity to the scale.

To summarize, the model tested shows good performance after some items that capture unfamiliar concepts are removed. It is important to note that, in such case, the conceptual model is being adjusted to the data. We would need to try to understand if there is some issue related to the validity of the items withdrawn and verify the adequacy of the questionnaire for the audience. We suggest that this issue be addressed in future research.

5.1.2 Conclusions for theory development

Lodahl e Kejner's (1965) scale had been criticized by authors such as Reeve and Smith (2001) and Saleh and Hosek (1976), because the item contents indicated that some concepts were not germane to the Work involvement construct. Due to that criticism, researchers discontinued its use; therefore, we cannot speak of the "destruction of theory" in this case, since this scale is not considered an established theory.

However, the scale was replicated in Brazil by Siqueira (2008). This author reduced the scale from 20 to 5 items. Regarding scale content, we can consider that Brazilian surveys ought to test the scales empirically, rather than adapting them after using exploratory techniques. We tested Lodahl and Kejner's (1965) model, which could not be confirmed with three dimensions – it was rejected on the grounds of being inadequate.

Lastly, from the point of view of the method of validation of scales – the focus of the present research -, the CFA technique proved to be the proper method for the validation of measurement scales.

5.2 Second study: Work values

Rokeach (1973) posits that human values are derived from the enduring beliefs that emanate from specific modes of conduct and from a final state of existence. Such values are chosen by the individual and compared to alternate behaviors to reflect a preferred mode of conduct. Thus, individual values can be viewed as concepts that carry a motivational connotation and determine general objectives that are applicable to a context.

The same author states that there are two categories of values: the terminal, and the instrumental values. The former are related to individual desires, whereas the latter represent the means chosen to achieve them. Individual choices depend on the factors, and sometimes lead to actions that are in opposition to the individual's personal values and beliefs. This generates conflict and psychic or emotional impact. In that sense, the more compatible the values are (with respect to the issues involving his or her work), the more intense will be the connection between satisfaction with work, productivity and personal development.

Along the same line, Schwartz (1992) posits that the values relate to individual attitudes and behaviors. The motivational values and their interrelationships are characterized by compatibility, antagonism and conflict. The author posits that there are conflicting values which lie in opposite directions, while there are other adjacent/compatible values. Schwartz's model is composed of ten categories, comprised of two main dimensions: (a) openness to change versus conservatism (in other words, values that emphasize autonomous thinking, as opposed to an emphasis on tradition, regulation, safety, conformity and security); and (b) values that emphasize the acceptance of others as equals and concern for general well-being (universalism and benevolence), as opposed to dominance of people (power and achievement).

Pasquali and Alves (2004) replicated Schwartz's (1992) Personal values scale which contains 40 items. Using Exploratory Factor Analysis, the authors suggest that the construct has four dimensions.

5.2.1 Empirical test of Pasquali and Alves's (2004) model

Our initial confirmatory test of the scale confirms the dimensionality of Pasquali and Alves's (2004) model. The adjustment indices found are acceptable, except for CFI – slightly lower than the recommended level of 0.9, according to Hair et al.'s (2009) recommendation. Still, the results lead us to accept the model (see Table 3).

Table 3 – Scale validation of Pasquali and Alves (2004)

Factors	Items	Model adjustment indices
F1	p12, p18, p19, p23, p27, p29, p34	normed χ^2 : 54
F2	p13, p17, p24, p32, p39	CFI: 0.889
F3	p28, p36, p38	RMSEA: 0.084 CI: 0.095
F4	p10, p26, p37, p11	SRMR: 0.074 GFI: 0.855

Source: Prepared by the author.

The Convergent Validity was subsequently analyzed (Table 4). The extracted mean variance and composite reliability are within the levels recommended by Hair et al. (2009).

Table 4 – Convergent validity of the PVQ-40 scale

Factors	VME	CC
F1	0.5228	0.8842
F2	0.5297	0.8484
F3	0.5182	0.7615
F4	0.5764	0.8442

Source: Prepared by the author.

To evaluate the model's discriminant validity, the χ^2 was calculated for the free model and for the alternative model, which has a fixed unit-value correlation between factors F1 and F2. The results depicted on Table 5 allow us to reject the null hypothesis that the correlation between factors is equal to one, meaning that the model has discriminant validity.

Table 5 – Test of χ^2 difference for scale discriminant validity

Factors	degrees of freedom	χ^2	Dif. χ^2	Probability ($>\chi^2$)	Sig.
F1, F2	53	71.418	26.893	2.15 ⁻⁷	0
	54	307.890			
F1, F3	34	93.838	93.705	2.205 ⁻³	0.001
	35	306.956			
F1, F4	43	21.359	17.388	3.047 ⁻⁵	0
	44	53.059			
F2, F3	19	17.154	14.722	1.246 ⁻⁴	0
	20	118.425			
F2, F4	26	23.561	23.236	1.433 ⁻⁶	0
	27	112.074			
F3, F4	13	5.059	15.068	1.037 ⁻⁴	0
	14	70.013			

Source: Prepared by the author.

To evaluate the criterion validity, a dependent variable related to the employee motivation was included in the structural model:

p41 - Do you feel motivated to perform positively in the work environment?

The empirical test of the model yields acceptable results (Table 6). This leads us to accept the hypothesis that an increase in the factors related to the work values (except for the fourth factor, which is not supported) leads to a moderate increase in the motivation to perform the job functions. This result fairly provides criterion validity for the scale.

Table 6 – Criterion validity for the Work values

Factors	Unstandardized loading	p(>z)	Standardized loading	R ²
F1	0.593	0.001	0.457	0.407
F2	0.246	0.011	0.189	
F3	0.250	0.070	0.192	
F4	-0.099	0.471	-0.076	

Source: Prepared by the author.

5.2.2 Conclusions for theory development and for measurement scale testing

The extant literature used exploratory techniques which are considered inadequate for the validation of measurement scales. In a different vein, the present research used confirmatory techniques to empirically test the Work values scale. In doing so, we identify items that do not contribute to measure the underlying concept. We also confirm the dimensionality of the construct, in line with the work of Pasquali and Alves (2004).

It is worth mentioning that – according to Muthén et al. (1997, apud Finney and Distefano, 2006, p. 294) - the minimum sample size for the WLSMV method is 200 cases. Despite this, the algorithm indicates that a sample of 204 is insufficient. This prompted us to use the traditional maximum likelihood method.

In this regard, the Soper sample size calculator (2017) considers that 200 cases are sufficient. But for non-normal or ordinal data, it suggests that the sample size be doubled. Thus, in future research the recommendation is to use of the WLSMV algorithm to replicate the scale, with a minimum sample of 400 cases.

6 DISCUSSION OF RESULTS AND CONCLUDING REMARKS

This article analyzed analytical techniques used in the construction of theory in the Business Administration field. As discussed, there are gaps in the application of statistical techniques for the construction of theory, due to the inadequate use of exploratory statistical methods. This issue usually results in contradictory findings against original theoretical frameworks.

This issue is quite relevant for theoretical research development in the field, since the use of exploratory techniques usually decomposes the previous theory. In addition, it becomes difficult to compare the results found with previous research, and in extreme cases there may be loss of the theoretical sense that the constructs provide.

This paper proposes a constructive research contribution, by way of a robust problematization that addresses the need to use the proper analytical procedures for validation of the research instruments.

This research's contribution for the field is given from the evidence provided by two studies. specifically: (i) the scale of Involvement with the work. After analysis, in line with other studies, the confirmatory test shows that the original instrument cannot be confirmed; we thus reject the model due to its inadequacy and; (ii) the model of Pasquali and Alves (2004), which was replicated and confirmed the construct's dimensionality.

Among the relevant contributions of this study is the categorization of research based on the degree of knowledge, as well as a proposed set of empirical test procedures that use confirmatory methods to develop theory in the field of Administration.

From the point of view of the validation of measurement scales, the research suggests proper courses of action that avoid situations where the original conceptual structure is not preserved. It also reinforces the use of confirmatory techniques to identify if the scale items capture "irrelevant spaces" of the construct, and their possible scale contamination.

From a common methodological perspective, the replication of scales illustrates the situations in which there are context differences, as well as problems of interpretation or questions related to the items validity. The reflections presented can help Academia reflect critically on procedures that are still prevalent.

7 RESEARCH LIMITATIONS AND SUGGESTIONS FOR FUTURE WORK

It is worth noting that Lodahl and Kejner's (1965) scale was criticized by a number of authors; yet, it served as the basis for Siqueira's "adapted" scale (2008). Still, the fact that it is not an established theory implies that it has not necessarily been "destroyed". This fact somehow limits the conclusions of the first study in terms of the research problematization.

A comparison of the original and Siqueira's adapted scale is not possible in this study, since data from the latter was not collected. Still, the empirical validation of the original scale was fully implemented. The caveat is that these findings can only be generalized for the researched population, and hence the original scale cannot be completely rejected.

There is a distinct possibility that the scale discrepancies arise from some other issue in the scale construction process – such as language translation, data collection procedures or common method bias. Still, the present study's findings are in line with the contributions made by Saleh and Hosek (1976) and Reeve and Smith (2001).

In the second study, a comparison of the original scales of Schwartz (1992) and a replication is not possible due to the use of a different analytical technique – namely, Schwartz did not use Factor Analysis, but multidimensional scaling. Therefore, the two study approaches are not fully balanced.

The fact that the empirical data from the original scale of Schwartz (1992) and Pasquali and Alves's (2004) "adapted" model cannot be compared limits the exploration of the determination of the "theory destruction".

In addition, note that the difficulties arise from an Organizational Behavior scales perspective from the fact that there is "some scale knowledge" (Figure 1). OB theories, in general, are still not considered to be established; we lack more empirical studies that focus on scale construction and validation. We conclude this work by suggesting that further research be conducted for the construction and consolidation of theories in Business Administration.

REFERENCES

- Armstrong, J., & Soelberg, P. (1967). *A Note on the Interpretation of Factor Analysis or Factor Analysis: What good is it?* Working Paper – Alfred P. Sloan School of Management –MIT.
- Bido, D. (2014). *Escalas como ferramentas de diagnóstico e gestão: Que peso dar aos dados (análise fatorial exploratória) e que peso dar à teoria e pesquisas anteriores (análise fatorial confirmatória)?* Apresentação oral no Painel Paralelo EPQ (P-EPQ4) - EnANPAD.
- Bido, D., Ribeiro, D., & Cohen, E. (2016). *Aplicações da Análise Fatorial Exploratória à pesquisa na área de Produção e Operações: Um estudo em publicações nacionais e internacionais*. Artigo apresentado no XIX Simpósio de Administração da Produção, Logística e Operações Internacionais, EAESP - Fundação Getulio Vargas.
- Borsa, J., Damásio, B., & Bandeira, D. (2012). Adaptação e validação de instrumentos psicológicos entre culturas: Algumas considerações. *Paidéia* (Ribeirão Preto), 22(53), 423-432.
- Brown, S. P. (1996). A meta-analysis and review of organizational research on job involvement. *Psychological Bulletin*. Washington, DC.
- Conway, J., & Huffcutt, A. (2003). A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational Research Methods*, 6(2), 147–168.
- Costa, F. (2011). *Mensuração e desenvolvimento de escalas: Aplicações em Administração*, Rio de Janeiro, RJ: Editora Ciência Moderna Ltda.
- Devellis, R. (2016). *Scale development: Theory and applications (applied social methods)* (4th ed.). Los Angeles: Sage Publications, Inc.
- Fabrigar, L., Wegener, D., Maccallum, R., & Strahan, E. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272.
- Finney, S., & Distefano, C. (2006). Non-normal and categorical data in structural equation modeling. In G. R. Hancock, & R. O. Mueller (Eds.). *Structural Equation Modeling: A second course* (p. 269–312). Greenwich: IAP - Information Age Publishing, Inc.
- Hair, J., Jr., Black, W., Babin, B., Anderson, R., & Tatham, R. (2009). *Análise multivariada de dados* (6a ed.) Porto Alegre: Bookman.
- Hair, J., Jr., Gabriel, M., & Patel, V. (2014). Amos Covariance-based structural equation modeling (CB-SEM): Guidelines on its application as a marketing research tool. *REMark – Revista Brasileira de Marketing*, Edição Especial, 13(2).

- Hinkin, T. (1995). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, 21(5), 967-988.
- Hinkin, T. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, 1(1), 104-121.
- Hinkin, T., Tracey, J., & Enz, C. (1997). *Scale construction: Developing reliable and valid measurement instruments* [Electronic version]. Retrieved from <http://scholarship.sha.cornell.edu/articles/613>
- Hunt, S. (1991). *Modern marketing theory: Critical issues in the Philosophy of Marketing Science*. Cincinnati, Ohio: South-Western Publishing Co.
- Jöreskog, K. (2007). Factor analysis and its extensions. In R. Cudeck, & R. MacCallum. (Ed.). *Factor Analysis at 100: Historical developments and future directions*. New Jersey: Lawrence Erlbaum Associates, Publishers, p. 47-77.
- Kline, R. (2010). *Principles and practice of structural equation modeling* (3th ed.). New York: Guilford Press.
- Lodahl, T., & Kejner, M. (1965). The definition and measurement of job involvement. *Journal of Applied Psychology*, 49, 24-33.
- Netemeyer, R., Bearden, W., & Sharma, S. (2003). *Scaling procedures: Issues and applications*. Thousand Oaks: Sage Publications.
- Osborne, J. (2014). *Best Practices in Exploratory Factor Analysis*. Louisville, NY: CreateSpace Independent Publishing Platform.
- Pasquali, L. (2012). *Análise Fatorial para pesquisadores*. Brasília, DF: LabPAM.
- Pasquali, L., & Alves, A. R. (2004). Validação do Portraits Questionnaire de Schwartz para o Brasil. *Avaliação Psicológica*, 3(2), 73-82.
- Pett, M., Lackey, N., & Sullivan, J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research*. Thousand Oaks: Sage.
- R Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Reeve, C., & Smith, C. (2001). Refining Lodahl and Kejner's job involvement scale with a Convergent Evidence Approach: Applying Multiple Methods to Multiple Samples. *Organizational Research Methods*, 4(2), 91-111, Sage Publications, Inc.
- Rokeach, M. (1973). *The nature of human values*. New York: Free Press.
- Rosseel Y. (2012). Lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1-36. Retrieved from <http://www.jstatsoft.org/v48/i02/>
- Saleh, S., & Hosek, J. (1976). Job involvement: Concepts and measurements. *Academy of Management Journal*, 19(2), 213-224.

- Schwartz, S. (1992). Universals in the content and structure of values: Theory and empirical tests in 20 countries. In M. Zanna (Ed.). *Advances in experimental social Psychology*, 25, 1-65. New York: Academic Press. Retrieved from [http://dx.doi.org/10.1016/S0065-2601\(08\)60281-6](http://dx.doi.org/10.1016/S0065-2601(08)60281-6)
- Siqueira, M. (2008). *Medidas do comportamento organizacional: Ferramentas de diagnóstico e de gestão*. Mirlene Maria Matias Siqueira (Org.). Porto Alegre: Artmed.
- Siqueira, M. (2014). (Ed.). *Novas medidas do Comportamento Organizacional: Ferramentas de diagnóstico e de gestão*. Porto Alegre: Artmed.
- Soper, D. S. (2017). *A-priori Sample Size Calculator for Structural Equation Models* [Software]. Recuperado de <http://www.danielsoper.com/statcalc>
- Worthington, R., & Whittaker, T. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, 34(6), 806–838.

Attachment A – Adjustment indices recommended from the literature

Test statistic	Measurement formula	Acceptable measurement levels
Chi-square estimate using maximum likelihood ratio (χ^2)	$(N - 1) \cdot \hat{F}$	The smaller the better $100 \leq N \leq 200$ (Hair Jr. et al., 2009)
χ normed	This metric tests the fundamental hypothesis of the model (measurement or structural), comparing the covariance matrix of the sample and the reproduced matrix; we do not want to refute the null hypothesis that there is no difference between the matrices	$p > 0.05$
Root of the standardized mean root square (SRMR)	q = number of variables in the analysis $SRMR = \left[2 \sum_{i=1}^q \sum_{j=1}^i \left(\frac{s_{ij} - \hat{\sigma}_{ij}}{s_{ii} \cdot s_{jj}} \right)^2 \cdot \frac{1}{q(q+1)} \right]^{1/2}$	SRMR < 0.10 acceptable SRMR < 0.05 ok model (Kline, 2010)
Goodness of fit index (GFI)	$GFI = 1 - \frac{F_{ML}}{F_0} = 1 - \frac{tr \left[\left(\hat{\Sigma}^{-1} \cdot S - I \right)^2 \right]}{tr \left[\left(\hat{\Sigma}^{-1} \cdot S \right)^2 \right]}$	The greater the better GFI > 0.9 = good fit (Kline, 2010)
Root mean square error of approximation (RMSEA)	$RMSEA = \sqrt{\frac{\max(\chi_M^2 - df_M, 0)}{df_M \cdot (N - 1)}}$	RMSEA < 0.05 good fit [between 0.05 and 0.08] ok > 0.08 Poor ≥ 0.1 unacceptable(problem) Higher threshold of the confidence interval ≤ 0.08
Comparative fit index (CFI)	$CFI = 1 - \frac{\max(\chi^2 - df_M, 0)}{\max(\chi_M^2 - df_M, \chi_{nulo}^2 - df_{nulo}, 0)}$	CFI ≥ 0.9