

TESTING AND VALIDATION OF A LEVELS OF CONSCIOUSNESS ASSESSMENT FOR LEADERS

Marianne Ales Stacheski¹

Luciana Gabardo²

Camila Marcon³

José Vicente B. de Mello Cordeiro⁴

ABSTRACT

Assessments of worldviews and levels of consciousness are critical for organizational interventions guided by an integral approach. Consultants and leaders might have a much easier task on promoting change if they can infer the levels of consciousness that are activated. Nevertheless, most of the assessments currently available present difficulties to effectively identify the process of thinking or the stage of consciousness one is really operating from. This work aims to develop, test and validate a new assessment to identify respondents' level of consciousness. It is a continuation of the previous research project, on which an internal validation with a Cronbach's Alpha coefficient of 0,8138 was performed, but there were still many issues to be addressed in order to reach an external validation. The new version of the assessment developed within this research presents two different sections, the first section contains thirty objective questions focusing on five different situations and the second section comprising sixteen subjective sentence completion questions (SCT). The new assessment was applied to 115 valid respondents but couldn't be completely validated due to methodological problems identified during the implementation process. However, two important fundamental conclusions for the future validation of the new assessment were reached. First, objective tests will more often access exposed values and not thinking processes. This aspect, once

¹ Aluna do 6º período do curso de Negócios Internacionais da FAE Centro Universitário. Bolsista do Programa de Apoio à Iniciação Científica (PAIC 2022-2023). *E-mail:* marianneless19@gmail.com

² Aluna do 6º período do curso de Psicologia da FAE Centro Universitário. Voluntária do Programa de Apoio à Iniciação Científica (PAIC 2022-2023). *E-mail:* lugbrd1903@gmail.com

³ Aluna do 3º período do curso de Administração da FAE Centro Universitário. Voluntária do Programa de Apoio à Iniciação Científica (PAIC 2022-2023). *E-mail:* camila.marcon@mail.fae.edu

⁴ Doutor em Engenharia de Produção pela Universidade Federal de Santa Catarina. Professor da FAE Centro Universitário. *Email:* josec@fae.edu

accepted, can be considered not only as a distortion of the test result, but can be used as a flag to generate hypotheses about the individual's level of consciousness, considering that the exposed values usually point to levels not developed or not fully developed yet by the individual. Second, levels of consciousness are not verifiable from contents but better from structures and from the whole answer set presented by an individual and not from the sum of each of their answers, which put the required analysis out of the reach of content analysis' softwares, demanding specific coding and automatizing.

Keywords: Assessment. Leadership Effectiveness. Level of Consciousness.

INTRODUCTION

As we face a period of crisis in leadership (BECK; COWAN, 1996; WILBER, 2000), it is urgent the need for leaders with more complex thinking processes, who have developed or are developing their worldview and consciousness level at the post-postmodern or integrative stage. Therefore, tools to identify thinking processes and consciousness levels, and to help leaders and their teams to overcome gaps and gain self-improvement, are also an urgent need.

Many tools were developed and have been used to assess leadership styles (ANDERSON; ADAMS, 2015; GOLEMAN, 2000) and adult development stages, which includes consciousness levels and worldviews (for example, LOEVINGER's WUSCT – Washington University Sentence Completion Test, 1979; KEGAN's SOI – Subject / Objective Interview, 1988; COOKE-GREUTER's SCTi-MAP – The Integral Sentence Completion Test, 1999; BECK; COWAN's SD – Spiral Dynamics Values Test, 1996). However, some of them demand a considerable fraction of time to be applied and extremely qualified professionals to grasp the results. Other tools, despite their recognized reliability, validity, and consistency, in the last years have been presenting some unsatisfactory results, assessing the subjects' shallower layer of values (exposed), which includes the values related to the contents of thoughts, not assessing thinking processes or the consciousness level in which their choices and actions are truly rooted. Such a scenario has been recurrently observed by one of the authors of this article (CORDEIRO et al., 2019) in his many years of teaching and serving as an integral leadership consultant.

On the last year, through the research program "PAIC – Programa de Apoio à Iniciação Científica" promoted by FAE – Centro Universitário in Curitiba (BRA), the authors developed a new tool aiming to assess more precisely the levels of consciousness and worldviews of leaders and their teams in general, also with less complex application and interpretation processes. The result was an assessment which encompasses two psychological tests combined: an objective or psychometric test (select-response), which provides initial references about the subject's level of consciousness, followed by a subjective test (sentence completion, free-response), which goes further and to confirm (or not) the initial references.

As a pilot study, the assessment was tested with 40 students from different programs and areas at FAE Business School, and its objective part had its internal validity and consistency confirmed by a 0,8138 Cronbach's Alpha coefficient. Furthermore, the subjective part of the assessment effectively allowed more accurate final conclusions about the thinking processes of the respondents in integration with the objective

answers. Nevertheless, especially in the objective test, the same main distortions in results presented by the existing tests were identified: the results were more about the subjects' shallower layer of values (exposed) and the contents of their thoughts, not their thinking processes or the consciousness level in which their choices and actions are truly rooted. Other aspects demonstrated the need for refinement or improvement in both parts of the assessment.

Considering the description above, the main research problem is summarised by one question: what are the necessary improvements in the original assessment to get more accurate results in mapping levels of consciousness and thinking processes? Since the purpose of this work is to achieve an external validation of the proposed assessment, it is necessary to reduce or even eliminate distortions, refining the assessment and repassing through a complete internal validation.

1 LEVELS OF CONSCIOUSNESS AND WORLDVIEWS THEORIES

This chapter is divided into the main theories about levels of consciousness and worldviews the Authors considered to develop the combined psychological tests (assessment) during the research program "PAIC – Programa de Apoio à Iniciação Científica – 2021/2022" promoted by FAE – Centro Universitário, in Curitiba (BRA), and in which they delved for the necessary refinement: Ken Wilber's Integral Theory, Don E. Beck and Christopher C. Cowan's Spiral Dynamics (SD), Suzanne Cook-Greuter's Ego Development Theory (EDT), and Robert Kegan's Development of the Self Theory.

1.1 THE INTEGRAL APPROACH

Ken Wilber's Integral Theory, developed as an evolution of the Developmental and Transpersonal approaches for Psychology, presents a quadrant model known as AQAL (All Quadrants, All Levels), representing the four basic-dimensions of reality: internal, external, individual, and collective.

TABLE 1 — Ken Wilber’s AQAL Model

AQUAL-Model - The Four Quadrants

UPPER LEFT I (experiential phenomena) Intentional (subjective)	UPPER RIGHT It (behavioural phenomena) Behavioural (objective)
LOWER LEFT We (cultural phenomena) Cultural (intersubjective)	LOWER RIGHT Its (social & systemic phenomena) Social (interobjective)

SOURCE: Esbjörn-Hargens (2010, adapted)

The upper left quadrant includes feelings, thoughts, beliefs, internalised values, worldviews and levels of consciousness. Cook-Greuter (2014) points that this quadrant encompasses ego development and the interior growth of self-awareness, including temporary states of consciousness. The upper right quadrant has an objective approach, emphasising on people’s behaviours, and the chemical and biological states in their bodies. The lower left quadrant represents the second-person and plural perspectives that one shares with the groups they relate to, focusing on the cultural aspect. Lastly, the lower right quadrant focuses on a third-person perspective, highlighting social and systemic overviews and environmental, political, educational, legal and economic issues, and also the structures, processes and activities within organizations (ESBJÖRN-HARGENS, 2010).

1.2 LEVELS OF CONSCIOUSNESS

Levels of consciousness’ concept, presented as systems of values including patterns of thinking processes were introduced by Clare W. Graves (1970). According to Graves, it is possible to note that man’s consciousness evolves from one level to another by quantum-like jumps in a progressive and spiral pattern. It is possible to achieve development when individuals can overcome and face new challenges that weren’t characteristics of their previous stages. Each level is equivalent to a set of knowledge that a society or individual uses for dealing with external stimuli.

Based on Gravesian theory, Don E. Beck and Christopher C. Cowan the Spiral Dynamics (SD), which classifies levels of existence into eight stages divided in two groups: warm colours (beige, red, orange and yellow) with a more individualistic bias and cool colours (purple, blue, green and turquoise) with a more communitarian perspective. Beck and Cowan (1996) presented that each colour can be encoded with a pair of letters, life conditions represented by “A” to “I” and “N” to “U” for ways of coping with reality.

In total harmony with Wilber's Integral Theory (2000) Beck and Cowan's Spiral Dynamics of levels of consciousness (1996) is presented as a probable sequence of personal evolution from biopsychosocial aspects, with the integration (and not exclusion) of the surpassed levels. According to the Spiral Dynamics, human consciousness evolves through a series of stages, each characterized by a different set of values, beliefs, and worldviews. The model uses a spiral metaphor to represent this evolution, with each stage building on the previous one.

The eight stages identified by Beck and Cowan (1996) are: 1) Beige: This stage is characterized by survival instincts and a focus on meeting basic needs such as food, shelter, and safety. 2) Purple: At this stage, people form strong bonds with their family and tribe, and are guided by superstitions and magical thinking. 3) Red: The Red stage is marked by a focus on power, domination, and individualism. People at this stage are driven by their own needs and desires, and seek to exert control over others. 4) Blue: At the Blue stage, people adopt a more traditional and hierarchical view of the world, based on moral absolutes and a belief in authority. 5) Orange: The Orange stage is characterized by a focus on individual achievement and rationality, and a rejection of traditional values and authority. 6) Green: At this stage, people emphasize equality, community, and diversity, and seek to create a more compassionate and inclusive society. 7) Yellow: The Yellow stage is marked by a focus on systems thinking and a willingness to consider multiple perspectives. People at this stage are able to transcend the limitations of previous stages and see the interconnectedness of all things. 8) Turquoise: At the Turquoise stage, people adopt a holistic and integrative worldview, and seek to create a sustainable and spiritually evolved society.

It's important to note that the Spiral Dynamics model is not intended to be prescriptive or judgmental. Rather, it is a tool for understanding the different ways that people see the world and make sense of their experiences. By recognizing the different stages of consciousness and values, we can better understand and communicate with people who hold different beliefs and perspectives.

Suzanne Cook-Greuter (1999; 2014) presented a nine-level model focused on maturity that deployed human capability towards a deeper comprehension and knowledge. The approach has a focus on sense-making based on three interrelated dimensions: operative or doing (goals), affective or being (emotions), and cognitive or thinking (about oneself and the world). Like the Spiral Dynamics' levels of existence, Cook-Greuter's phases alternate between integration (phases 3, 4, 5, and 6) and differentiation (phases 2/3, 3/4, 4/5, and 5/6)

The chart below briefly presents Cook-Greuter’s Ego Development Theory stages, excluding Symbiotic Phase 1, which is not covered in the maturity chart:

TABLE 2 — The Nine Stages of increasing embrace in Ego Development Theory

Tier	Stage	Name	Self	Others	Main characteristics
Preconventional	2	Impulsive Stage	Rudimentary, physical self-labelling (me, mine)	Objects that exist to fulfil needs on demand	Constant sense of threat, insecurity; governed by impulses; limited understanding of the world
	2/3	Self-Protective or Opportunistic	Single concrete features	Having their own needs and desires: competition	Manipulation; no use of arguments and logic; intimidation; low trust; others are always to blame.
Conventional	3	Conformist or Diplomat	Several external features; rudimentary internal states	Two types: my group (family, tribe, nation) and the out-group	Group-centric thinking, self-identity depends on the group, belonging, group values, shared truths, strong morality (good or bad), traditions.
	3/4	Self-Conscious or Expert	Separated from others; sense of specialness	Separated and different. Need for constant comparison	Beginning of self-understanding; focused energy; need to be different / special; sense of knowing-it-all (superiority).
	4	Conscientious or Achiever	In society; drive for achievements	With similar goals and aspirations	Full awareness of linear time (reflections about past, present, and future); need for broader relational social context; need for achievements; rational and determined.
Postconventional	4/5	Individualist or Pluralist	A participant-observer, fourth person perspective	Equally valid and worthy of consideration	Recognition of multiple perspectives, awareness of the complexity of systems, vivid individualism, all ideas are respected, relativism.
	5	Autonomous or strategist	Embedded in history and multiple cultural contexts	Embraced for what they are	Historical perspective (lifetime), multiple contexts (time and space), well balanced, insightful, integration of conflicting aspects. Meaning making.
Postautonomous	5/6	Magician ou Alchemist	Beyond own lifetime; global-historic perspective	To be served; we all need each other	Inner conflicts about existential paradoxes, accepting paradoxes and ambiguity, committed to serving self and others;
	6	Unitive	Witnessing, not observing	We are all part of oneness. No ego boundaries	Meaning making with compassion; no illusions about permanence; belongingness and separateness without tension.

SOURCE: Cook-Greuter (1999; 2014, adapted)

The “Impulsive” level: At this level, individuals are driven by their impulses and immediate needs. They have little capacity for self-reflection or perspective-taking, and are often focused on getting their own needs met. The “Self-Protective” level: At this level, individuals become more aware of their own needs and desires, but are still primarily focused on self-preservation and self-defense. They may be skeptical of others and struggle with trust and vulnerability.

The “Conformist” level: At this level, individuals begin to identify with and conform to social norms and expectations. They may rely on external validation and approval to feel secure, and may struggle to think for themselves. The “Expert” level: At this level, individuals become more self-aware and reflective. They may begin to question their assumptions and beliefs, and develop a more nuanced understanding of themselves and others. The “Conscientious” level: At this level, individuals become more conscientious and responsible, and may feel a sense of duty to others or to a higher cause. They may develop a strong sense of ethics and morality, and may feel a deep sense of responsibility for their actions.

The “Individualistic” level: At this level, individuals begin to prioritize their own individuality and autonomy, and may seek out experiences and relationships that allow them to express themselves freely. They may also begin to question authority and social norms, and develop a more critical perspective. The “Autonomous” level: At this level, individuals become more self-directed and independent. They may feel a sense of inner freedom and autonomy, and may be less attached to external validation or approval. They may also begin to see the world in a more holistic and interconnected way.

The “Integrated” level: At this level, individuals become more integrated and balanced in their thinking and behavior. They may have a more nuanced understanding of themselves and others, and may be able to hold multiple perspectives at once. They may also feel a sense of peace and contentment, and be less driven by external goals or desires. The “Unitive” level: At this level, individuals experience a sense of oneness and interconnectedness with all things. They may feel a deep sense of love and compassion for all beings, and may be less attached to their own individuality or identity. They may also experience a sense of transcendence or spirituality.

Kegan’s perspective focuses not so much on the phases, but much more on the process, the developmental dynamics in each phase and between phases, considering that people spend most of their lives in transitions. On the author’s reasoning, the subject / object distinction is a form of understanding what will be experienced through a process of differentiation and integration, the capability to emerge from rootedness in something (like values, perceptions, feelings) taking it out of our subjectivity and turning it into a new object, something to reason about, understand and integrate. This process “is essentially that of adaptation, a differentiation from that which was the subject of my personal organization and which becomes thereby the object of a new organization” (KEGAN, 1982, p. 75-76).

The table below presents a summary about Kegan’s stages:

TABLE 3 — Robert Kegan’s Stages of Development of the Self

Stage	Name	Perspective	Subject	Object
1	Impulsive	Single perspective, can’t take other’s point of view	Immediate perceptions, feelings, and impulses	Actions, sensations, and physical objects
2	Instrumental/ Imperial	Can take multiple perspectives, but one at a time	Enduring interests, personal agendas, & role expectations	Immediate perceptions, feelings, and impulses
3	Interpersonal/ Socialised	Can take two or more perspectives simultaneously	Shared meaning, mutuality, social ideals & self-consciousness	Enduring interests, personal agendas, & role expectations
4	Institutional/ Self-authoring	Own personal perspective on relationships and societal ideals	A self-authored system of values and standards	Shared meaning, mutuality, social ideals & self-consciousness
5	Interindividual / Self-transforming	Recognizes that own perspective on experience is a self-created convenience	Universality, paradox, multiple “selves” as vehicles for connection	A self-authored system of values and standards

SOURCE: Kegan (1982, adapted)

At the impulsive stage (1), the child is tightly integrated on the only perspective (immediate perceptions, impulses and desires), being unable to weigh everything. In the imperial stage (2), all the embedded aspects become enduring interests. Kegan claims that most adults function at stage 3, interpersonal or socialized stage, for they have the ability to reason from different perspectives and points of view, and also understand relationships with shared expectations and meanings. The theory explains that at stage 4 the individual is no longer formed by relationships and ideals of society, but only possesses them. There is a true self, self-sufficiency, self-regulation and completeness, whatever the intimate relationship and context. Finally, in stage 5, Interindividual / Self-transforming, there is a self that creates other identities in each setting and relationships, and this process becomes the subject of the individual. His values and models are continually changing, as they are no longer the self, but the self he chooses to be and change according to the context.

Moreover, it is relevant to note some central points that all these theories have in common. First, the levels and stages presented do not imply any moral supremacy, only a more evolved sense of maturity (REITTER, 2014) or more elaborate ways of reasoning, in other words, doing and understanding. These theories intend to offer tools to understand the thought processes that one has when facing the world everyday. Also considers that human development occurs in a logical and ongoing process that relies on the support and challenge of the environment, “from simple to complex, from static to dynamic, and from ego-centric to socio-centric to world-centric” (COOK-GREUTER, 2014, p. 2).

Lastly, the theories are also interconnected in that previous phases or levels must be integrated, learning and evolving to the next level, none of which can be ignored. It is true that some obsessions over previous stages occur: they exist and indicate that some issues are not fully resolved. And to move completely from one to another, these issues must be overcome, so the contests of each phase must be fully understood and integrated, not dismissed or ignored.

2 REVIEWING FORMER ASSESSMENT

The former assessment encompassed two psychological tests combined: an objective or psychometric test (select-response), which provides initial references about the subject's level of consciousness, followed by a subjective test (sentence completion, free-response), which goes further and to confirm (or not) the initial references.

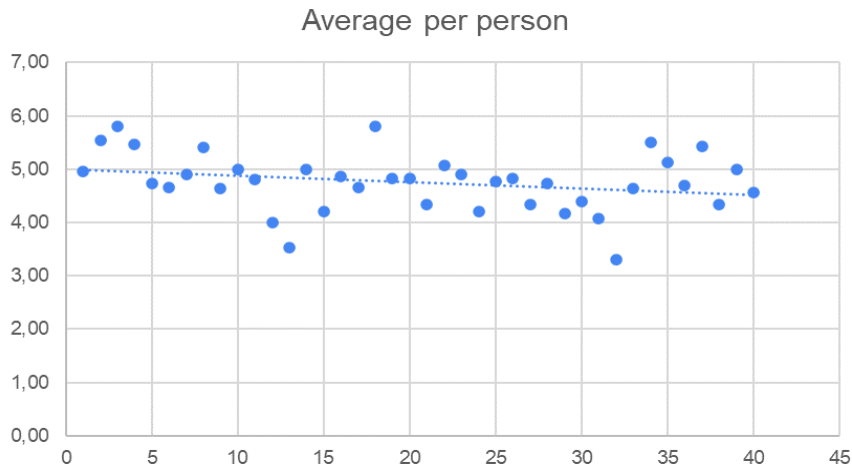
Through the process of internal validation of the former assessment, the authors identified some distortions and aspects which needed refinement.

About the distortions in the objective or psychometric test (select response), many respondents showed positive identification with affirmations of the Spiral Dynamic Red level of existence, a much lesser complex and more surpassed level that, in general, dominates the thinking and acting process in times of tension and conflict, especially in people who operate mostly in the SD orange level. This result indicates a probable failure in the test construction, especially the choice of words for the red responses.

Sentences to identify transition from Blue (Conformist) to Orange level (Conscientious), that is, Cook-Greuter's Expert level, was not formulated in the previous questionnaire. However, since a significant amount of resultants were operating in the SD orange level, they could actually be in a transition, when analysing in parallel the SCT test. Regarding the absence of questions highlighting the Expert level and the distortion it can have caused, the authors judged it important to formulate them for the new test.

Second, the results presented an average line on the "Agree, but sometimes I am not like this" response (5,00, Chart 1), which aims to identify values and idealizations (usually a later stage of development), with a tendency to "Neither disagree nor agree" (4,00, Chart 1). This result showed an acquiescence bias phenomenon, when people tend to agree with the statements proposed and at the same time it showed a normal distribution, when people tend to be in the middle, neither agreeing nor disagreeing.

CHART 1 — Average of the Likert Scale's answers from the respondents



SOURCE: The authors (2022)

Third and at last, the general results in the objective part of the assessment showed the same main distortion the authors intended to overcome: they map more the exposed values than the thinking processes of the respondents, which was confirmed through the integration with the sentence completion test (subjective, free-response) results.

About the sentence completion test (the subjective part of the assessment), although it has achieved the purpose of mapping more the thinking processes of the respondents than their exposed values, it was not free from the need for refinement. Many respondents gave extremely short answers (even with one word) which were useful only in the total context of the results, demanding enormous work of interpretation by the applicator. It became evident the need for sentences that lead the respondents to express more of their thoughts, concerns, and feelings.

3 METHODOLOGY

The authors began this research with the analysis of the former assessment results in its interval validation to identify distortions and aspects in need for refinement, which were pointed out in chapter 1.

In the regard of the objective test, considering the inadequacy of some affirmations (especially in the red level ones), the absence of affirmations for the Expert level, and the general main distortion (accessing exposed values, not thinking processes), the authors decided to remake the whole path to construct a new test, following Domino and Domino (2006) and Urbina (2014).

Regarding step 2 (theoretical framework), authors revisited three levels of consciousness and worldview theories: Beck and Cowan's Spiral Dynamics (SD), Cook-Greuter's Ego Development Theory, and Kegan's Development of the Self Theory. Differently from the former assessment construction process, the authors emphasized the two former theories to build the new affirmations for the objective test, considering they are both more focused in thinking processes and are freer from level contents than the SD Theory.

About step 3, the authors explored psychological tests construction's rules and theories, especially regarding scales. It is mandatory to identify a scale which provides less space to acquiescence bias phenomenon, leading the respondents to more personal and specific choices.

The same approach (remaking the construction steps) was done regarding the subjective test, with one big difference: the authors intend to maintain the same themes elected for the former proposal, just giving to the sentences to be completed new structures to lead respondents to more complex answers. For that, it was necessary further understanding about the process of Robert Kegan's SOI – Subject / Object Interview, based in his Development of Self Theory, designed to explore and identify the complexity of the subject's meaning-making and status on the individual evolutionary journey. The intention was not only to get better results in mapping the respondents' thinking processes but also making the results interpretation task simpler.

It is important to point out the necessity of revisiting the mentioned levels of consciousness and worldview theories not only to remake the steps for refining the assessment, but also to identify possible causes of the issues and distortions verified in the former proposal and overcome them. For example, in that regard, it might help to get further knowledge in the differences between exposed values, contents and thinking processes, and to understand more accurately the process of fluctuation of the levels of consciousness operating in one person, according to different situations.

In constructing the questionnaire, one topic which deserves attention is the attitude scale such as the number of choices or points on the Likert scale (CHANG, 1994). There are several discussions when it comes to choosing the number of points. According to Haladyna and Rodriguez (2013), the effect of decreasing the number of options proportionally increases test efficiency for large numbers of respondents and decreases efficiency for small numbers of respondents. Since the authors' main goal is to achieve an internal and external validation of the assessment as a complete study, the number of respondents is going to be expressively higher. Another consideration is that more options are usually more indicated when respondents master the subject

under study or when the object of study has many attributes (DALMORO; VIEIRA, 2013). According to this work's objective and taking it all into account, it was chosen to reduce the number of points and to refine the nominal answers in identifying with the proposed statement.

Changes were not only about reducing the number of points but also making an objective test where the respondent has the opportunity to choose between the items of each question, the one or the ones with which they identify the most. And also being able to classify all the items according to its identification as "weak", "moderate", "strong" or "very strong". The statements proposed in the former psychological test passed through a deep revision and reformulation by the authors.

Finally, it is important to mention that this is a qualitative research since it is supposed to explain a phenomenon, that is the effective mapping of the levels of consciousness, further than numerical data. It is an applied study which aims to solve the distortion in identifying the actual thinking process someone is operating in. The modality of this research, besides of having a bibliographic and descriptive character, is mixed that is both qualitative and quantitative. Since our product is an instrument capable of collecting data, survey research methods are applied in this work. The reason which justify this modality is the convergent delineation characterised by qualitative and quantitative data collection and analyses during the same time and stage of the process. It is followed by a fusion of both types of data in a general interpretation (CRESSWELL; CLARK, 2013). The purpose of this modality is to obtain different data but complementary to better understand the research's problem (GIL, 2021).

The exploratory literature review on the main theories aims to sustain the research's instrument. Such as Cook-Greuter (2014) and Robert Kegan (1982; 1998), main references for this work. Regarding the need for a refinement of the previous instrument, exploratory revision was adopted in our methodology to increase the references about construction, administration and validation of an assessment. Besides the review, a deep analysis and comparison of the previous instrument took place in the rebuilding of this work's instrument.

The authors intended the validation process to be divided into three core parts: internal validation, internal consistency, and external validation.

The Internal validation of an assessment refers to the process of evaluating the reliability and validity of the instrument that is being used to measure a construct of interest. The purpose of internal validation is to ensure that the assessment is measuring what it is intended to measure (SOUZA; ALEXANDRE; GUIRARDELLO, 2017), and that the results are consistent and reliable over time.

There are several methods of internal validation (SULLIVAN, 2011), including: 1) Test-retest reliability, which involves administering the same assessment to the same group of individuals at two different points in time, and comparing the scores to determine if they are consistent over time; 2) Internal consistency, an evaluation of the extent to which different items on the assessment measure the same construct. This can be done through methods such as Cronbach's alpha or split-half reliability, which will be later explained in the topic *Internal Consistency*; 3) Inter-rater reliability, which consists in assessing the degree of agreement between different raters who are administering the same assessment; 4) Factor analysis, an analysis of the underlying structure of the assessment to determine if the items are measuring distinct factors or dimensions.

One important topic when it comes to validation is the size of the sample, the minimal number of the sample must be five times bigger than the number of items to be evaluated (HAIR JR. et al., 2005). The authors aimed to have a reliability index (Alpha) in 95% and a 5% error margin, using a simulator of sample the sample size must be in 384 respondents. There is no mandatory requirement for sample size, but a sample of over 300 or at least 10 observations per item of the draft scale is recommended to achieve a stable result (FENN; CHEE-SENG; GEORGE, 2020).

About the Internal Consistency, Cronbach's alpha is a measure of internal consistency reliability that is commonly used to assess the reliability of assessments. It is a coefficient that ranges from 0 to 1, with higher values indicating greater internal consistency among the items or questions in an assessment. To apply Cronbach's alpha to an assessment, first it is needed to administer the assessment to a group of individuals and obtain their responses. Then, to calculate the correlation between each item or question and the total score on the assessment, as well as the correlation between each item or question and the other items or questions in the assessment. The resulting alpha coefficient ranges from 0 to 1, with higher values indicating greater internal consistency among the items or questions in the assessment.

The external validation of an assessment refers to the process of evaluating the performance of an assessment tool by comparing its results to external criteria or measures (SOUZA; ALEXANDRE; GUIRARDELLO, 2017). The purpose of external validation is to determine the accuracy, reliability, and validity of the assessment tool.

One common method of external validation is to compare the results of the assessment to the results of other measures that are known to be valid and reliable. For example, if an assessment tool is designed to measure reading comprehension, its results can be compared to the results of other reading comprehension tests to determine if they are consistent. Another method of external validation is to compare

the results of the assessment to other measures of the same construct, such as other tests that measure the same skill or ability. This approach can help to establish the convergent validity of the assessment, which means that it measures the same thing as other established measures (DOWNING, 2003).

Finally, external validation can also involve comparing the results of the assessment to real-world outcomes or criteria that are relevant to the construct being measured (COOK; HATALA, 2016). For example, if an assessment tool is designed to measure leadership performance, its results can be compared to actual leadership performance ratings to determine if they are predictive of success as a leader.

There is evidence to suggest that an interview can serve as a form of external validation of assessments. In a study by Konrad, Wiek, and Barth (2019), researchers found that interviews were a valuable tool for validating assessments of soft skills, such as teamwork and communication. The study found that interviews provided an opportunity for candidates to demonstrate their soft skills in a way that was not always captured by assessments. In addition, the use of structured interviews, where all candidates are asked the same questions in the same order, has been shown to improve the validity and reliability of the interview process (MCDANIEL et al., 1994).

4 RESULTS: NEW ASSESSMENT DESIGN, TESTING AND VALIDATION

The new (or second version) of the proposed assessment consists of two phases: objective questions and subjective statements.

The objective test consists of a questionnaire composed of twenty questions with seven items based mainly on Cook-Greuter (2014) and Kegan (1982; 1998) theories, but also on having insights from Beck and Cowan's Spiral Dynamics (1996) and Anderson and Adams' Mastering Leadership approach (2015). A review of the previous assessment was conducted as well to refine but mainly to develop items with less distortions.

The twenty questions were inspired on the upper left side of Ken Wilber's Integral Theory AQAL model, presenting topics such as attitude, inner motive, interactions, fear, trust, communication. They also contained values, resulting in questions such as "I would like", "I wouldn't like", "I admire", "I reject", despite the initial intentions to eliminate them. One important conclusion this work has made is that using exposed values as allies can produce a more precise diagnostic than to concentrate on eliminating them. The authors arrived at this conclusion by realizing that exposed values are majorly representing a posterior hot/cold level from where the individual operates at.

For example, one operating on Orange (hot colour) commonly identifies itself with or appreciates Yellow values, the next hot colour. Meanwhile one operating on Blue (cold colour) tends to prefer Green aspects, skipping the Orange level. When a result consists of a high Blue and a high Green but a rejected Orange it means Orange level wasn't fully integrated thus automatically one can be operating on Cook-Greuter Expert level, a transition between SD's Blue and Orange levels. This is a common result observed on the actual and on previous assessments.

Regarding the items, theories presented on this work were consumed and reviewed to their construction. The definition for each level was used in the construction process. By Red representing conflict; Blue needs; Orange work; Green convictions/values; Yellow leadership; Turquoise relationship with the self. A couple of verbs presented on the definitions and also on the Leadership Circle Assessment (ANDERSON; ADAMS, 2015) are presented on a couple of items.

Stages of Development (KEGAN, 1982) represents the increased complexity of thought while the self evolves on age and consciousness. One moves from having simple perceptions to sustaining an intrapersonal relation, from enduring dispositions to inter-personalism, from abilities of self-regulation, authorship, and formation to a dialectical cognition inter-institutional and a transformational interpenetration of selves (KEGAN, 1998). Kegan's theories and terms served as inspiration for the development of questions, especially those corresponding to more complex processes of thought.

Nine Levels Of Increasing Embrace In Ego Development (COOK-GREUTER, 2014) forms the base for construction of the assessment. It was substituted for Beck & Cowan's Spiral Dynamics (SD) due to the formerly identified importance of the transition between SD's Blue and Orange levels. So, each answering item was added an alternative corresponding to Cook-Greuter's Expert level. An explicit Expert representation on the assessment became extremely necessary since the transition from SD's Blue to Orange are the levels on which most of educated adults operate from in our society nowadays (CORDEIRO; CRUS FILHO; LAMOGLIA, 2019). The definition of Cook-Greuter's nine levels has given a robust understanding about each of them. Language clues and expressions common of each level such as "I win, you lose", "social glue", "I am special, I deserve respect", "make rational and equitable decisions", "who am I? why do I act like that?", "purpose", "connected and alone" (COOK-GREUTER, 2014) contributed for the development of items even more specific to the respective levels.

The new version of the subjective part of the assessment (the sentence completion test) is composed of sixteen sentences to be completed, designed mostly according to Robert Kegan's procedure to administer the "subject-object" interview (2011). The sentences referred to contents presented by Kegan as "rich areas" or

“flags” that in general lead the respondents to more reflection: responsibility, guilt, failure, disappointment, expectations, personal and social contradictions, reciprocity, consideration, competition and motivations.

The new structure given to the sentences to be completed effectively gave rise to larger responses, although some answers continued short (in some cases, extremely short, with one or two words).

Following Kegan’s procedure, before analyzing the answers for the sentence completion test, the authors considered the objective test results to generate a hypothesis or a range of hypotheses about the individuals underlying subject-object structures and levels of consciousness. And through the analysis of the completed sentences, the authors observed how the contents were organized, from which perspective the answers were expressed, and the depth of the understanding (what is subject and what is object for the respondents). This research aimed, initially, to automatize SCT answers with some content analysis software, like ATLAS TI. Nevertheless, it was identified that the coding enabled by this sort of software couldn’t consider the answers for different items at the same time. Considering that mental processes cannot be identified by one word or expression within the answer to an item, but only by the context on which they are given, authors decided to run an one on one analysis for SCT answers, planning a further automatization of the test with the help of artificial intelligence.

The new assessment was tested with 115 respondents: 25 students from FAE Business School MBA’s Leading Yourself class, 36 from FAE Business School Graduate Specialization’s Integral Leadership class, and 54 employees from MPS Informática, all located in Curitiba (BRA).

The SCT – sentence completion test results from the 61 students were analysed by the authors and confirmed by one of them through activities made during the two different graduate/MBA classes. The levels recognized with the SCT results were confirmed in those activities.

The results from the 54 employees from MPS Informática were internally confirmed by crossing the objective results with the SCT ones, analysed by the authors.

On answering the assessment the respondent is supposed to select the item or items with which have most resonated, one was allowed to select one to three items for each question. That means the Likert scale was eliminated, in order to reduce a previously identified bias on the initial assessment, that is, the tendency for identification at least partially with most of the sentences, which had led to a difficulty to distinguish the centre of gravity of many of the respondents of the former assessment. Authors concluded that choosing between statements and opening for more than one selection in each item could be a form to get closer to the respondent’s mind structure and perhaps to catch

one's transition from one level to another. However, eliminating the scale brought a few new difficulties in measuring the assessment. Google Forms' lack of capability to handle different weights among the one to three alternatives selected on each item contributed to some lack of conclusiveness regarding respondents' level of consciousness. Giving too much freedom resulted in respondents being not very specific when selecting the items, some selected more than three while a few focused on picking up strictly the essential ones. Since there was no attitude scale the authors evaluated according to the number of items selected for each question or for each level resulting in a final number, which was expected to be the one correspondent to the respondent centre of gravity.

The elimination of the Likert Scale made unfeasible a new internal validation with Cronbach's Alpha. This happened mostly because of the possibility to select multiple items out of a common scale and the unthought regarding a possible punctuation. Such as a respondent who selected fewer items has a higher score than those who selected more. When it is taken on account that the first one put some extra thought while answering the assessment, but it was not a topic which received enough studies.

Technical issues had also accompanied the assessment's application on the chosen platform (Google Forms) for the questionnaire. It wasn't possible to limitate the number of items one could select, leading to a confusion when analysing the whole data. Even though the instructions had the minimum and maximum number of answers it is possible to observe that some respondents didn't follow the instructions, ending up invalidating their answers. Regarding the respondents, the minimum sample capable of satisfying the criteria established by authors wasn't achieved, which resulted in an incomplete pilot study. Hence neither an internal nor an external validation were correctly accomplished due to the lack of respondents.

Besides the undone and unfinished phases mandatory for complete consistency and validation, positive conclusions were reached with this research. In applying the assessment authors concluded that distortions presented in the previous assessment were solved in the new one. And, most importantly the product achieved its main goal, to map one's level of consciousness more precisely, focusing on mental structures and processes and not on exposed values. When evaluating the obtained results, the authors concluded a paramount method for an analysis to be effective. The objective test has to be a pre-analysis which has to be confirmed or not with the subjective one (SCT). When analysing both objective and subjective parts of the 115 respondents from MPS Informática Ltda. and FAE Business School respondents, empirically the results matched. Again it is important to mention that this empirically analysis was made by one of the authors who is sufficiently qualified to judge accordingly. Even though the experiments were made with small samples in a manually one on one analysis, the assessment has shown its effectiveness accomplished its purpose.

FINAL REMARKS

Most of the world's most renowned authors in the field of psychological testing agree that the development of a psychological test can take many years, and may have to be rephrased and repeated several times (ANASTASI, 1954; DOMINO; DOMINO, 2006; URBINA, 2014). This paper's authors now have a complete understanding of this reality.

The necessary enhancement of the first version of the proposed assessment was successful in correcting some distortions, and presented adequate results: most of the hypotheses generated by the objective test results were confirmed by the sentence completion test results. The analysis of the SCT answers, manually made by the authors, showed that its results were very consistent with the objective questions and enabled the proper identification of deep internalized values and thinking processes.

However, by producing the new version of the proposed assessment, and by analyzing not only its results, but also the first version results and some other existing assessment results, the authors finally deeply understood two aspects of paramount importance.

First, objective tests, due to their fundamental structure, more often will access exposed values and not thinking processes. This aspect, once accepted, can be considered not only as a distortion of the test result, but can be used as a flag to generate hypotheses about the individual's level of consciousness, considering that the exposed values usually point to levels not developed or not fully developed yet by the individual.

Second, one same content can be processed in different ways according to the individuals' levels of consciousness. In other words, the levels of consciousness are not verifiable from contents but better from structures and from the answer set presented by the individuals. That being so, the Atlas TI software, capable enough to do content analysis using predefined terms and expressions, is not the most suitable tool to be applied for the assessment automation purposes.

A possible topic for future research could be the development of a suitable software for the complete automation of the assessment. For this, it would be necessary to analyse the given answers to the sentence completion test, group them according to the levels of consciousness indicated, and from this analysis build objective statements which better express the structures of the levels of consciousness. From the research done so far, it appears that the best assessment structure to be automated would require groups of objective statements, the first one involving all levels to create hypotheses, and the subsequent groups to eliminate hypotheses, with questions of less and less levels until the final result is reached.

REFERENCES

- ACEVEDO, A. A Personalistic Appraisal of Maslow's Needs Theory of Motivation: From "Humanistic" Psychology to Integral Humanism. **Journal of Business Ethics**, New York, v. 148, n. 4, p. 741-763, Apr. 2018. <https://doi.org/10.1007/s10551-015-2970-0>
- ANASTASI, A. **Psychological testing**. New York: The Macmillan Company, 1954.
- ANDERSON, R. J.; ADAMS, W. A. **Mastering leadership: an integrated framework for breakthrough performance and extraordinary business results**. New York: Wiley, 2015.
- BECK, D.; COWAN, C. **Spiral dynamics: mastering values, leadership and change**. London: Blackwell Publishing, 1996.
- BROWN, T. Construct Validity: A Unitary Concept for Occupational Therapy Assessment and Measurement. **Hong Kong Journal of Occupational Therapy**, v. 20, n. 1, p. 30-42, Jun. 2010. [https://doi.org/10.1016/s1569-1861\(10\)70056-5](https://doi.org/10.1016/s1569-1861(10)70056-5)
- CATTELL, R. B. **The scientific analysis of personality**. 2. ed. New York: Routledge, 2017.
- CHANG, L. A Psychometric Evaluation of 4-Point and 6-Point Likert-Type Scales in Relation to Reliability and Validity. **Applied Psychological Measurement**, v. 18, n. 3, Sept. 1994. <https://doi.org/10.1177/014662169401800302>
- COOK, D. A.; HATALA, R. Validation of Educational Assessments: A Primer for Simulation and Beyond. **Advances in Simulation**, v. 1, n. 1, Jan. 2016. <https://doi.org/10.1186/s41077-016-0033-y>
- COOK-GREUTER, S. **Ego development: a full spectrum theory of vertical growth and meaning**. 2014. 122 p. Available on: https://www.researchgate.net/publication/356357233_Ego_Development_A_Full-Spectrum_Theory_Of_Vertical_Growth_And_Meaning_Making. Accessed on: July 3rd 2022.
- COOK-GREUTER, S. **Postautonomous ego development: its nature and measurement**. Cambridge, MA: Harvard Graduate School of Education, 1999.
- CORDEIRO, J. V. B. M. **A dinâmica da espiral e o problema do Brasil**. **Integral Works**, 10 set. 2016. Disponível em: <https://integralworks.com.br/noticias/iw16002>. Acesso em: 09 out. 2023.
- CORDEIRO, J. V. B. M.; CRUZ FILHO, P. R. A.; LAMOGLIA, L. B. **Liderança Integral: a evolução do ser humano e das organizações**. Petrópolis: Vozes, 2019.
- CRESWELL, J. W.; CLARK, V. L. **Pesquisa de métodos mistos**. 2. ed. Porto Alegre: Penso, 2013.
- DALMORO, M.; VIEIRA, K. M. Dilemas na construção de escalas tipo Likert: o número de itens e a disposição influenciam nos resultados? **Revista Gestão Organizacional**, v. 6, ed. spe., p. 161-174, 2013.
- DAVIS, J. An overview of transpersonal psychology. **The Humanistic Psychologist**, Washington, v. 31, n. 2-3, p. 6-21, Apr. 2003.
- DOMINO, G.; DOMINO, M. L. **Psychological testing: an introduction**. 2. ed. Cambridge: Cambridge University, 2006.
- DOWNING, S. M. Validity: on the meaningful interpretation of assessment data. **Medical Education**, v. 37, p. 830-837, 2003.

ESBJÖRN-HARGENS, S. An overview of integral theory: an all-inclusive framework for the 21st Century. In: ESBJÖRN-HARGENS, S. (Ed.). **Integral theory in action**: applied, theoretical, and constructive perspectives on the AQAL model. New York: State University of New York, 2010. p. 33-61.

FENN, J.; CHEE-SENG, T.; GEORGE, S. Development, validation and translation of psychological tests. **BJPsych Advances**, v. 26, n. 5, p. 306-315, 10 Jun. 2020. <https://doi.org/10.1192/bja.2020.33>

FINN, S. E. et al. Psychological testing and psychological assessment: a review of evidence and issues. **American Psychologist**, Washington, v. 56, n. 2, p. 128-165, Mar. 2001. doi.org/10.1037/0003-066x.56.2.128

GIL, A. C. **Como fazer pesquisa qualitativa**. Barueri: Atlas, 2021.

GOLEMAN, D. Leadership that gets results. **Harvard Business Review**, v. 78, n. 2, p. 78–90, 2000.

GRAVES, C. W. **Levels of existence**: an open system theory of values. New York: Union College, Schenectady, 1970.

HAIR JR., J. F. et al. **Análise multivariada de dados**. 5. ed. Porto Alegre: Bookman, 2005.

HAIR, J. et al. **Multivariate data analysis**. 4. ed. Englewood Cliffs: Prentice Hall, 1995.

HALADYNA, T. M.; RODRIGUEZ, M. C. **Developing and validating test items**. New York: Routledge, 2013. Available on: [books.google.com.br/books?hl=pt-BR&lr=&id=-uL5BQPFO3MC&oi=fnd&pg=PP2&dq=MC+Rodriguez+\(2005\)](https://books.google.com.br/books?hl=pt-BR&lr=&id=-uL5BQPFO3MC&oi=fnd&pg=PP2&dq=MC+Rodriguez+(2005)). Accessed on: Aug. 1st 2023.

HAYNES, S. N.; RICHARD, D. C. S.; KUBANY, E. S. Content validity in psychological assessment: a functional approach to concepts and methods. **Psychol. Assess.**, v. 7, n. 3, p. 238-247, 1995. doi.org/10.1037/1040-3590.7.3.238

INGLEHART, R.; WELZEL, C. **Modernization, cultural change, and democracy**: the human development sequence. United Kingdom: Cambridge University Press, 2005. p. 57.

KEGAN, R. **In over our heads**: the mental demands of modern life. Boston: Harvard University, 1998.

KEGAN, R. **The evolving self**: problem and process in human development. Boston: Harvard University, 1982.

KEGAN, R. et al. **A guide to the subject-object interview**: its administration and interpretation. Raymond: Minds at Work, 2011.

KLEINEBERG, M. **Integrative levels**. Toronto: Knowledge Organization, 2017.

KONRAD, T.; WIEK, A.; BARTH, M. Embracing Conflicts for Interpersonal Competence Development in Project-Based Sustainability Courses. **International Journal of Sustainability in Higher Education**, v. 21, n. 1, p. 76-96, 6 Jan. 2020. <https://doi.org/10.1108/ijshe-06-2019-0190>

KUBANY, E. **Content validity in psychological assessment**: a functional approach to concepts and methods. **Psychological Assessment**, v. 7, n. 3, p. 238-247, Sept. 1995. Available on: https://www.researchgate.net/publication/232480869_Content_Validity_in_Psychological_Assessment_A_Functional_Approach_to_Concepts_and_Methods. Accessed on: Oct. 9th 2023.

MCDANIEL, M. A. et al. The validity of employment interviews: a comprehensive review and meta-analysis. **Journal of Applied Psychology**, v. 79, n. 4, p. 599-616, 1994. <https://doi.org/10.1037//0021-9010.79.4.599>

OOSTROM, J. K., et al. Why do situational interviews predict performance? Is it saying how you would behave or knowing how you should behave? **Journal of Business and Psychology**, v. 31, n. 2, p. 279-291, 9 Jun. 2015. <https://doi.org/10.1007/s10869-015-9410-0>

REITTER, N. Clare W. Graves and the turn of our times. **Journal of Conscious Evolution**, San Francisco, v. 11, n. 11, p. 1-66, Nov. 2014.

ROGERS, C. The necessary and sufficient conditions of therapeutic personality change. **Psychotherapy: Theory, Research, Practice, Training**, Washington, v. 44, n. 3, p. 240-248, Oct. 2007. doi.org/10.1037/0033-3204.44.3.240

SOUZA, A. C.; ALEXANDRE, N. M. C.; GUIRARDELLO, E. B. Psychometric properties in instruments evaluation of reliability and validity. **Epidemiologia e Serviços de Saúde**, v. 26, n. 3, p. 649-659, Jul. 2017. <https://doi.org/10.5123/s1679-49742017000300022>

SULLIVAN, G. M. A Primer on the validity of assessment instruments. **Journal of Graduate Medical Education**, v. 3, n. 2, p. 119-120, June 2011. <https://doi.org/10.4300/jgme-d-11-00075.1>

TORBERT, W. R. **Brief comparison of five developmental measures**: the WUSCT, the SOI, the LDP, the MAP and the GLP primarily in terms of pragmatic and transformational validity and efficacy. 2014. Available on: <http://www.williamrtorbert.com/wp-content/uploads/2012/03/DevMeasuresCompared4714.pdf>. Accessed on: Sept. 30th 2022.

TRANFIELD, D. et al. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. **British Journal of Management**, v. 14, n. 3, p. 207-222, Sept. 2003.

URBINA, S. **Essential of psychological testing**. 2. ed. New Jersey: Wiley, 2014.

VALLE, R.; HARARI, C. Transpersonal psychology. **The Humanistic Psychologist**, Washington, v. 13, n. 1, p. 11-15, Mar. 1985.

VICH, M. The origins and growth of transpersonal psychology. **Journal of Humanistic Psychology**, Thousand Oaks, v. 30, n. 2, p. 47-50, Aug. 1990.

WILBER, K. **Integral psychology**: consciousness, spirit, psychology, therapy. Boston: Shambhala, 2000.

WILBER, K. **The collected works of Ken Wilber**. Boston: Shambhala, 1999-2000.

WILBER, K. **The integral vision**. Boston: Shambhala, 2008.

WITT, A. et al. A new tool to map the major worldviews in the Netherlands and USA, and explore how they relate to climate change. **Environmental Science & Policy**, Amsterdam, v. 63, p. 101-112, Sept. 2016.