

# Construct validity of the Organic ScoreCard<sup>©</sup>

Dr NW Simpson

## Abstract

Although classical models on validity divided validation into various "validities" (such as content validity, criterion validity and construct validity), the current view is that validity is a single construct. Using Messick's unified and expanded theory of validity, which includes the evidential and consequential bases of test interpretation, this paper evaluates the validity of the Organic ScoreCard<sup>©</sup> (OSC). The Organic ScoreCard<sup>©</sup> is a 108-item scale measuring the concept of "consciousness". After addressing each of the key aspects of construct validity (consequential, content, substantive, structural, external and generalisability), it was concluded that the Organic ScoreCard<sup>©</sup> has construct validity on each of these aspects.

Keywords: consequential validity, content validity, empirical research, external validity, generalisability validity, structural validity, substantive validity, test interpretation, test validity, validity evidence

## Introduction

The Organic ScoreCard<sup>©</sup> is a 108-item scale measuring the concept of "consciousness". It was developed by Marcus Andreas Grond, a Dutch philosopher, business consultant and behavioural coach (Grond, 2005). The scale is being used commercially for coaching and consulting purposes within different industries in the Netherlands, South Africa and other parts of the world. It is a tool for developing self-consciousness and insight in persons functioning in different work environments. It is also used as an assessment tool and basis for interventions to promote healthy workplace behaviour. In addition, it is used as a personal coaching and therapeutic tool. The Organic ScoreCard<sup>©</sup> questionnaire is completed by respondents in their own time and the findings are presented as an automated rapport. Customised feedback sessions can be arranged with individual clients or groups of clients (typically workplace groupings like teams or departments).

Interpretation is derived from the resulting score combinations. The scores are plotted on a radar diagram representing the three broad groups of brain energy in red, blue and yellow areas. The significance of each pattern is then discussed.

Clients also receive a booklet which outlines the various aspects of the assessment tool's findings. The booklet serves as a guideline to help facilitate an understanding of brain-based strengths and weaknesses in various domains.

Grond did empirical research on thousands of Organic ScoreCard® participants in over 20 countries over a period of ten years, showing that detailed and personalised profiling is possible and that this information could serve as a basis of valuable insight and well-targeted coaching and mentoring interventions for measurable personal and professional growth.

Looking at behaviour, Grond did not employ comparison to others (peer groups) to gain an understanding of a client's "make-up". Instead, he searched for the 'why' of people's actions in the evolutionary driving forces of the brain. This gives each individual full credit for the life-long building of a strategy to deal with life and the people in it. The findings of the Organic ScoreCard® therefore provides insight into the 'how' (behaviour) and 'why' (origin of life strategy) of the individual, as well as the individual's future strategy in order to bring about more inner guidance, authenticity and happiness (the inner committed next step).

To structure the findings of the Organic ScoreCard® and to make the findings practical and applicable for each individual, a total of 36 different archetypes were defined. People get motivated by their specific combinations of these 36 'inner longings', as Grond calls them. These longings underpin people's strategies.

Grond distinguishes seven levels of depth in human awareness. The automated Organic ScoreCard® reports only provide feedback on the first three levels. The full value and depth of the Organic ScoreCard® (levels 4 to 7) is available through the direct feedback of certified Organic Coaches or Organic Therapists.

## New conceptualisation of construct validity

The concept of validity was traditionally defined as "the degree to which a test measures what it claims, or purports, to be measuring" (Brown, 1996, p. 231). Construct validity is thus the demonstration that a test is measuring the construct it claims to be measuring (Brown, 2000).

Construct validity can be studied in various ways and can be demonstrated from a number of perspectives. Hence, the more strategies used to demonstrate the validity of a test, the more confidence test users have in the construct validity of that test. In short, the construct validity of a test should be demonstrated by the accumulation of evidence.

A test is valid to the extent in which it accurately measures what it is supposed to measure (Brown, 2000). Validity therefore refers to the degree to which evidence and theory support the interpretations of test scores. Although classical models divided validation into various "validities" (such as content validity, criterion validity and construct validity), the current view is that validity is a single construct.

According to Trochim (2006), construct validity is the overarching quality with all of the other validity measurement labels falling under it: “When we claim construct validity, we're essentially claiming that our observed pattern – how things operate in reality – corresponds with our theoretical pattern – how we think the world works. I call this process pattern matching, and I believe that it is the heart of construct validity.”

Messick (1989) presented a unified and expanded theory of validity, which included the evidential and consequential bases of test interpretation and use. Table 1 shows how this theory works. Notice that the evidential basis for validity includes both test score interpretations and test score use. The evidential basis for interpreting tests involves the empirical study of construct validity, which is defined by Messick as the theoretical context of implied relationships to other constructs. The evidential basis for using tests involves the empirical investigation of both construct validity and relevance/utility, which are defined as the theoretical contexts of implied applicability and usefulness.

Table 1: Aspects of test validity according to Messick

	<b>Test interpretation</b>	<b>Test use</b>
Evidential basis	Construct validity	Construct validity + relevance and utility evidential basis
Consequential basis	Value implications	Social consequences

The consequential basis of validity involves both test score interpretations and test score use. The consequential basis for interpreting tests requires making judgements of the value implications, which are defined as the contexts of implied relationships to good/bad, desirable/undesirable, etc. score interpretations. The consequential basis for using tests involves making judgements of social consequences, which are defined as the value contexts of implied consequences of test use and the tangible effects of actually applying that test (Brown, 2000).

In 1989, Messick presented a new conceptualisation of construct validity as a unified and multi-faceted concept. His unified theory was the culmination of debate and discussion within the scientific community over the preceding decades. For example, Hughes (2018) argued that "Many of these sources of 'validity evidence' are not just difficult to combine, they are sometimes diametrically opposed and cannot be meaningfully represented within a unified model". Focusing on seminal works on the concept of validity over the last century, Messick commented on the “surprisingly tumultuous state of validity theory and practice within the twenty-first century”.

Messick's unified theory of construct validity covered six aspects of construct validity:

- Consequential
- Content
- Substantive
- Structural
- External
- Generalisability.

In the next section, these aspects will be applied to the Organic ScoreCard<sup>®</sup> to help determine construct validity.

## Types of validity-supporting evidence

### 1. Consequential

*Core question: What are the potential risks if the scores of the Organic ScoreCard<sup>®</sup> are invalid or inappropriately interpreted?*

According to Shepard (1997), "To study the consequences of a test is as much significant as the test's internal features such as validity and reliability".

Consequential validation has a proposed interpretation that begins with the framework that defines the scope and aspects of the multidimensional scales.

The Organic ScoreCard<sup>®</sup> is an easy-to-use tool to incorporate insights from awareness theory into our way of working and living. The Organic ScoreCard's automated reports leave no room for misinterpretation in understanding the outcome of any Organic ScoreCard<sup>®</sup>. There is thus no direct risks in interpreting the OSC.

The Organic ScoreCard's published guidelines used in training in several countries clarify the interpretation even further for certified coaches. The interpretation is thus supported by:

- Automated reports
- The certification of coaches
- Handbook and training manuals
- The transferability of theory in train-the-trainer programmes.

*Consequential validity:* The Organic ScoreCard's automated reports as well as handbooks and training of coaches minimise the potential risks that the scores of the Organic ScoreCard<sup>®</sup> can be invalid or inappropriately interpreted.

## 2. Content

*Core question: Do test items in the Organic ScoreCard® appear to be measuring the construct of 'awareness'?*

Haynes, Richard and Kubany (1995) described content validity as “the degree to which elements of an assessment instrument are relevant to and representative of the targeted construct for a particular assessment purpose”.

Content validity includes any validity strategies that focus on the content of the test. To demonstrate content validity, testers investigate the degree to which a test is a representative sample of the content of whatever objectives or specifications the test was originally designed to measure (Brown, 2000).

By dividing “awareness” into awareness of individuality, time and space, and subdividing each of these items into four aspects (inside/outside – I / We; yesterday/tomorrow – perception/reality; surroundings/world – tangible/intangible), 12 domains of awareness were defined. These 12 domains cover all angles of awareness (and action). Each of these domains can be viewing the world from the perspective of ‘survival’ (brainstem orientation), ‘connection’ (limbic orientation) or ‘trust’ (neocortex orientation). This gives a total of 36 awareness differentiations.

The Organic ScoreCard® measures affinity on these 36 items. As shown in Organic ScoreCard® tests all over the world, these 36 are recognised and embraced by everyone, differing in individual scores, but cross-culturally adequate and complete (Grond, 2005).

An important consequence of this ‘root-cause analysis’ approach of individual behaviour is that it makes every behaviour ‘logic’ and understandable (though not agreeable per se). It takes the ‘good’ and ‘bad’ out of the equation. Understanding the *why* and showing the *how* gives way to ‘root-impacted improvement’, inner growth and redevelopment, ultimately (and sometimes very quickly) leading to new behaviour. This new behaviour is grounded, not on outside demands, but on authenticity and the need for personal growth.

Since the first Organic ScoreCard® was rolled out in 2003, multiple redefinitions, adaptations and refinements of the questionnaire followed, with several groups of practitioners, coaches and teachers participating in the process. The current version of the Organic ScoreCard® is the twelfth version (2012).

According to test subjects worldwide, the outcome of the Organic ScoreCard® is accurate and consistent. Exceptions are rare and questionable.

This is how the questionnaire’s stability was brought about:

- 2003: The first prototype of the Organic ScoreCard® was an open-question handwritten query.

- 2004: The test items were standardised. The original formulation of the 108 test items was automated to serve as an online assessment tool.
- 2012: Collaborating with a group of high school teachers, the definitions were finalised.
- 2011-2014: Multiple translations followed, helping to prove construct stability.
- 2011-2018: The Organic ScoreCard<sup>®</sup> was tested by bilingual participants to help ensure common understanding and reliable results.

In 2004, the assessment tool's test items were standardised based on a sample of approximately 150 test participants. In 2018, 12 000 test persons have been registered in a reassessment process. Since 2012, the new test items standards were implemented, covering the period from 2012 to 2018. The Organic ScoreCard<sup>®</sup> has now been tested by over 12 000 people from 20 countries (Western Europe, Eastern Europe, Far East, South Africa (various ethnic backgrounds), Northern Africa, Maoris, American Indian).

*Content validity:* The elements of the Organic ScoreCard<sup>®</sup> assessment instrument are relevant to and representative of the construct of “awareness”.

### 3. Substantive

*Core question: Is the theoretical foundation underlying the construct of ‘awareness’ sound?*

One of the advantages of the Organic ScoreCard<sup>®</sup> is the way in which it looks at awareness. The tool presents awareness in a way that makes the unconscious workings of the brain visible and accessible.

The Organic ScoreCard<sup>®</sup> therefore allows for a unique “inside-out” view of individuals – solely based on their own and ultimately private history of building a strategy in life. It is about what is really driving people to action or lethargy. Once this is known, people can be guided to make changes leading to a more fulfilled and meaningful life.

Validation begins with a framework that defines the scope and aspects (in the case of multi-dimensional scales) of the proposed interpretation. Stone (2003) said that validity is “a theory guiding its development, a hierarchy of illustrative items constructed to define the variable, the subsequent production of item difficulties and person measures, and the analysis of fit.”

The 36 archetypes for organic awareness that the Organic ScoreCard<sup>®</sup> identifies are the result of comprehensive research on organic awareness (Grond, 2005).

The theoretical foundation underlying the Organic ScoreCard<sup>®</sup> is explained in various publications, including:

- Grond, M.A. (2005). *Organisatie burnout*. Kerk Avezaath NL: Management Book International.

- Grond, M.A. (2017). *Organisch Handboek*. Arnhem NL: TransMind International.
- TransMind South Africa. (2018). *Organic Manual*. Cape Town.
- Grond, M.A. (2018). *A model of goal integration*. Heerlen NL: Open University.

### **Underlying theories**

The Organic ScoreCard<sup>®</sup> is based on several underlying theories as described in *The Organisation Burnout* (Grond, 2005) and *The Organic Handbook* (Grond, 2017) as well as in the unpublished research paper by Grond titled: “A model of society/individual goal integration in education, based on the comparison between governmental demands, (classical) pedagogical theory and neuro-psychology, and its effects on daily school practice”.

### **Inter-correlation in an ipsative scale**

The Organic ScoreCard<sup>®</sup> is an example of an ipsative scale, which is a scale that presents the respondent with blocks of items that have to be rank-ordered. The Organic ScoreCard<sup>®</sup> presents items in 12 blocks of 9 items each.

Ipsative scales are different to other scales in that a linear dependency is built into their score structure. Score sums are the same for all responses and any scale score is linearly dependent on the other scale scores. Thus, even if different scores are derived for "reptile", "limbic" and "Neocortex" functioning, the final score will always be 120. Thus, different to the Likert scale that relies on variance in scores, the Ipsative scale does not rely on variance because there is no variance in the scoring mechanism. Respondents simply arrange their preferences into different quantities or proportions of the "trait" being measured. This option introduces local dependency or inter-correlation.

The common scoring mechanism in ipsative scales introduces a linear dependency among items. Hence, the correlations between ipsative items and external criteria cannot be independent. This complicates the interpretation of scale scores. Systems such as opposites on a continuum or the use of a spider diagram to show opposite scores are sometimes used to facilitate score plotting (Matthews & Oddy, 1997). Ipsative scales do not lend themselves to either factor analysis or reliability analysis as found in scale development methodology.

*Substantive validity:* The 36 archetypes for organic awareness identified by the Organic ScoreCard<sup>®</sup> is underpinned by solid theory, a hierarchy of illustrative items constructed to define the variables and fitting subsequent items for measurement.

## **4. Structural**

*Core question: Do the interrelationships of dimensions measured by the Organic ScoreCard<sup>®</sup> correlate with the construct of ‘awareness’ and test scores?*

Matthews and Oddy (1997) said that, “Structural validity gives ‘greater recognition ... to the validity and reliability of observed scores from measurement instruments. Specifically, measurement error has become a major issue in many disciplines...”

Structural validity has to do with the interrelationships of the structural components of the assessment instrument. This interrelationship is evaluated by looking at the structural relations of items or subscale interrelationships and scoring models and their consistency with what is known about the construct of “awareness”. This is often done using free sort tasks or factor analysis. For structural validity this evaluation should support the multidimensional conceptualisation of the Organic ScoreCard®. Structural validity is thus about the structural component of validity, the domains and scoring model of the construct of “awareness”.

The questions of the Organic ScoreCard® focus on core affinities (that you may or may not be aware of). This creates a picture of who you really are at that specific point in time, and what your opportunities and capacities are.

Awareness is divided into 12 ‘windows’ through which one can look at the world. In Organic Theory these windows are called domains. We have a brain strategy and it plays out in our behaviour. The Organic ScoreCard® makes our brain strategy visible in the 12 domains. Domains and their combinations are cross-cuts of our awareness. These cross-cuts are controlled by our brain in three different ways: Survival, Connection and Trust.

Each statement in the Organic ScoreCard® relates directly to one of the 12 items of this construct. Nine statements are directly and specifically linked to one of these 12 items.

- Automated link between statements and items
- Parallel systems used for every country and language.

*Structural validity:* The Organic ScoreCard® has strong structural relations between items and subscales and the scoring models and is consistent with the construct of “awareness”. The structural component (domains and scoring model) of the instrument thus supports the multidimensional conceptualisation of the construct of “awareness”.

## 5. External

*Core question: Does the test have convergent, discriminant and predictive qualities?*

According to Calder, Phillips and Tybout (1983), the question of external validity has to do with “whether the results of a behavioral study would hold for other persons, settings, times, or places”. They argued that the concept of external validity is less important than other forms of validity when the objective of research is to test theory. Their position is that external validity is “a matter of the applicability of behavioral research ... 'real world' variables ... only become important in the context of evaluating interventions based on theory”.

Lynch (1999) argued that external validity can only be assessed by “better understanding how the focal variables in one’s theory interact with moderator variables that are seen as irrelevant

early in a research stream. Findings from single real-world settings and specific sets of 'real' people are no more likely to generalize than are findings from single laboratory settings with student subjects. Both the laboratory and real world vary in background facets of subject characteristics, setting, context, relevant "history", and time. It is only when these facets vary and we see how they interact that understanding of external validity is enhanced."

The Organic ScoreCard<sup>®</sup> has been used in change and development programmes for individuals, teams and organisations. It has also been used in single measurements, correlation measurements and re-measurements after interventions. Its potential to guide processes, create groups (of interest, drive and so on) and differentiate between people (in employee recruitment/selection) is remarkably predictive. It has applications in cultural change processes, scholarly educational and career coaching.

- Proven applicability in individual coaching and therapy, in individual career guidance and self-development
- Proven applicability in team coaching and change guidance
- Proven applicability in organisational change processes
- Single and correlation measurements over time as well as over different groups
- Re-measurements after interventions (pre and post measurements, and gap analysis)
- Division markers on awareness items, group interventions based on that.

*External validity:* The Organic ScoreCard<sup>®</sup> uses findings from a variety of theoretical and real-world facets of subject characteristics, settings, contexts, histories and time to evaluate how they interact. This enhanced the understanding of external validity.

## 6. Generalisability

*Core question: Does the Organic ScoreCard<sup>®</sup> assessment generalise across different groups, settings and tasks?*

Generalisability theory is about how reliable observations are. It is therefore used to determine the reliability of measurements under specific conditions. According to Williams and Vaske (2003), in the use of psychological tests, generalisation facets often pertain to the test-taking conditions (e.g. location, occasions, proctors). According to them, "A good measurement instrument should minimize variance arising from these facets". In most scale development efforts (like in the Organic ScoreCard<sup>®</sup>) there is thus only one differentiation facet (i.e. one object of measurement).

In this context, Leung (2015) explained that generalisability is not expected of most qualitative research studies. They are meant to study a specific issue in a certain population or group, of a

focused locality in a particular context. However, he said, “with the rising trend of knowledge synthesis from qualitative research ... (the) evaluation of generalizability becomes pertinent”.

What are standards/norms? In order to be able to say something about the test values of a single person, the values of this person cannot be compared and related to a large norm sample. The Organic Awareness Theory presupposes that each individual has a unique combination of archetypes and that these change over time.

The 12 domains of awareness in three brain regions make 36 archetypes. This gives us a comprehensive translation of organic awareness into human behaviour. These archetypes tell the stories of success, pitfalls, dreams and things to do.

The 36 archetypes for organic awareness differ from behaviour-based lists in psychology because they were developed from theories of consciousness and then applied to thousands of people. Other approaches are often based on pure introspection, observations of behaviour, or in-depth psychological approaches.

The 36 archetypes:

- Have a deep integrated theoretical basis
- Were empirically validated
- Are based on field evaluations
- Have been confirmed across cultures.

These 36 archetypes are independent "factors" that have a high explanatory value in relation to human awareness. The archetypes also have a high predictability of behaviour. Every person has a unique awareness in the now.

The Organic ScoreCard<sup>®</sup> has proven itself as a useful and reliable tool in diverse contexts:

- Over 12 000 Organic ScoreCards<sup>®</sup> administered
- in over 20 countries
- to a diversity of people (aged 13 to 87; high school to PhD; entry-level employees to executives)

*Generalisability validity:* The Organic ScoreCard<sup>®</sup> has proven generalisability validity. It has been administered to a wide range of people in countries all over the world, proving to be a useful and reliable tool across a range test-taking conditions.

## Application

Research done by Dr Frederick Marais (advisor) and Ankia Du Plooy at the Unit for Innovation and Transformation (Ekklesia: Stellenbosch University) concluded that the Organic ScoreCard<sup>®</sup> used in conjunction with training and coaching interventions yielded measurable results.

The following themes emerged from the thematic content analysis of the one-on-one interviews with participants in the Organic ScoreCard<sup>®</sup>-supported intervention applied in an IT environment:

- **Personal growth through increased self-awareness and group awareness:** In terms of awareness, the intervention made a significant positive impact on all the respondents. The biggest impact on a personal level was that respondents had gained insight in terms of self-awareness (“understand myself better”) and group awareness (“understand the team better”; “appreciating the different gifts of different people”; “without a doubt changed my perspective”). This led to increased team cohesion and team functioning (“helped to work more collaboratively”; “were able to make new connections”), better communication (“helped to go straight to the point”) and increased outputs, which are crucial in agile environments. In essence, this helped participants to acquire emotional maturity.
- **Increased clarity on team roles, leading to increased confidence and empowerment of team members:** Team members felt empowered in the functional roles, which created confidence that impacted personal lives as well as team functioning (“team roles were affirmed and that created more confidence”; “good understanding of the team and how they work together to become more productive”; “helped working with different teams and knowing the unique angle where to work from as well as knowing the unique energy of the different teams”). The intervention created a space in which people were able to undergo “huge shifts” in their lives. This also helped them to cope with the trauma of organisational change.
- **Increased team functioning and team outputs:** Team members agreed that the intervention led to significantly enhanced team functioning (“less judgement in the team”; “took away the barriers and guards”; “a lot more tolerant towards the other team members”; “all are different, but that the differences complement each other”; “closer relationships improved the team’s functioning”; “much more collaboration and respectful interaction”; “not only black and white, also pay attention to the emotional side of people”; “the biggest team function change would definitely be understanding”; “it just made me feel more sure of what I was doing in my team, and the contribution I was making to the team”).
- **Stronger leadership and management skills:** Respondents in managerial positions commented that the tools and skills they had acquired during the intervention helped them to be better managers and leaders (“gave leaders tools to enhance their leadership

style and to use in managing and creating the teams”; “helped from a management perspective to understand how to engage with different individuals and different teams”).

- **Ideal fit with agile methodologies in IT environments:** The intervention supported the ability of team leaders to put together high-performing and self-sufficient teams (“better communication enhanced performance and individual discipline”; “could form teams to fit each other and complement each other to improved performance”; “for agile methodologies to work you really need to understand the people in the team that you’re working with, as well as possible, because it requires a lot more interaction between individuals than another process”; “and one of the things agile strives to do is to take the personal individual qualities, and take those into account in the way teams function. And the intervention certainly supported that and helped the team to be able to use the individual traits they had better, and to understand each other better, which is very central to agile”; “build trust in the teams, which is something from the agile environment you really try to build from the beginning”).
- **Enhanced ability to deal with change:** A significant number of participants made mention of their new approach to change and their ability to handle change better after the intervention. The intervention was particularly valuable to a number of IT people working for a company in the process of restructuring (“a company going through restructuring and retrenchments is like a slow motion car crash... and the intervention gave first aid”; “handled change more maturely and used it as a growth”; “because they went through deep change on a personal level that positively impacted them, to be in a better position to handle the organisational change”).
- **Efficiency of the intervention:** The intervention used a combination of one-on-one coaching sessions (“appreciated the deep listening”; “one-on-ones deepened the whole journey”) and facilitated workshops to enhance performance. Prior to the actual intervention, participants were invited to do the Organic ScoreCard<sup>®</sup>. This was well received by the participants (“you could drive straight to the point and didn’t spend months sitting in therapy until you eventually stumbled on it. So it was really a tool to give you a map, and then through discussion and so on it facilitated us getting to the issue really quickly... it saved a lot of time, the Organic ScoreCard<sup>®</sup> (helped you) to go straight where we actually needed to work”; “I was surprised that out of so few questions you could get such an accurate picture”). Participants also commented on the integrated approach to performance enhancement (“was not only about work, but an integrated process about life”; “helped to evolve as a person”; “positive towards the company who made the investment”; “this was by far the most insightful exercise I think I’ve ever done”).

In essence, the Organic ScoreCard<sup>®</sup>-supported intervention helped participants to acquire the emotional maturity and confidence to optimise their outputs in fast-moving agile environments.

## Conclusion

The value of the Organic ScoreCard<sup>®</sup> is underpinned by the growing traction of this assessment tool, ongoing research and its construct validity. The findings of the thematic content analysis of the one-on-one interviews with participants in an Organic ScoreCard<sup>®</sup>-supported intervention applied in an IT environment support the applicability of the instrument.

## References

- Brown J.D. (2000). Shiken: JALT Testing & Evaluation. *SIG Newsletter*, 4(2), 8-12.
- Calder, B.J., Phillips, L.W., & Tybout, A.M. (1983). The concept of external validity. *Journal of Consumer Research*, 10(1), 112-114.
- Grond, M.A. (2005). *Organisatie burnout*. Kerk Avezaath NL: Management Book International.
- Haynes S.N., Richard, D. & Kubany, ES. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, 7(3), 238-247.  
<http://dx.doi.org/10.1037/1040-3590.7.3.238>
- Hughes, D.J. (2018). Psychometric Validity: Establishing the Accuracy and Appropriateness of Psychometric Measures. In Irwing, P., Booth, T., & Hughes, D.J. (Eds.), *The Wiley Handbook of Psychometric Testing: A Multidisciplinary Reference on Survey, Scale and Test Development*. Hoboken: New Jersey: Wiley-Blackwell.
- Leung, L. (2015). Validity, reliability, and generalizability in qualitative research. *Journal of Family Medicine and Primary Care*.
- Lynch, J.G. (1999). Theory and external validity. *Journal of the Academy of Marketing Science*, 27(3), 367-376.
- Matthews, G. & Oddy, K. (1997). Ipsative and normative scales in adjectival measurement of personality: problems of bias and discrepancy. *International Journal of Selection and Assessment*, 5(3), 169-182.
- May, L.A., & Warren, S. (2002). Measuring quality of life of persons with spinal cord injury: external and structural validity. *Spinal Cord*, 40, 341-350.
- Messick, S. (1989). Meaning and Values in Test Validation: The Science and Ethics of Assessment. *Educational Researcher*, 18(2), 5-11.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessment. *Educational Researcher*, 23(2), 13-23.
- Messick, S. (1996). *Validity and washback in language testing*. Research Report, Educational Testing Service (ERIC Document Reproduction Service No. ED403277).
- Popham, W.J. (1997). Consequential validity: Right Concern – Wrong Concept. *Educational Measurement: Issues and Practice*, 16(2), 9-13.
- Shepard, L.A. (1997). The centrality of test use and consequences for test validity. *Educational Measurement: Issues and Practice*, 16(2), 5-8.
- Stone, M.H. (2003). Substantive scale construction. *Journal of Applied Measurement*, 4(3), 282-297.

Trochim, W.M.K. (2006). *Reliability*. The Research Methods Knowledge Base. Retrieved from socialresearchmethods.net

Williams, D.R. & Vaske, J.J. (2003). The measurement of place attachment: Validity and generalizability of a psychometric approach. *Forest Science*, 49(6). Retrieved from [https://www.fs.fed.us/rm/value/docs/psychometric\\_place\\_attachment\\_measurement.pdf](https://www.fs.fed.us/rm/value/docs/psychometric_place_attachment_measurement.pdf)