Choosing an Appropriate Data Collection Instrument and Checking for the Calibration, Validity, and Reliability of Data Collection Instrument Before Collecting the Data During Ph.D. Program in India

H. R. Ganesha [1] & Aithal P. S. [2]

[1] Research Professor, Institute of Management & Commerce, Srinivas University, Mangaluru, India, and Board Member, Gramss Retail Trading Private Limited, Bengaluru, India,
OrcidID: 0000-0002-5878-8844; E-mail: hrganesha@yahoo.co.in

[2] Professor & Vice-Chancellor, Srinivas University, Mangaluru, India,
OrcidID: 0000-0002-4691-8736; E-mail: psaithal@gmail.com

Area/Section: Research Methodology.
Type of the Paper: Conceptual Study.
Type of Review: Peer Reviewed as per |C|O|P|E| guidance.
Indexed in: OpenAIRE.
DOI: https://doi.org/10.5281/zenodo.7310078
Google Scholar Citation: IJMTS

How to Cite this Paper:
Ganesha, H. R., & Aithal, P. S., (2022). Choosing an Appropriate Data Collection Instrument and Checking for the Calibration, Validity, and Reliability of Data Collection Instrument Before Collecting the Data During Ph.D. Program in India. *International Journal of Management, Technology, and Social Sciences (IJMTS)*, *7*(2), 497-513. DOI: https://doi.org/10.5281/zenodo.7310078

International Journal of Management, Technology, and Social Sciences (IJMTS)
A Refereed International Journal of Srinivas University, India.

CrossRef DOI: https://doi.org/10.47992/IJMTS.2581.6012.0235

Received on: 10/10/2022
Published on: 10/11/2022

Choosing an Appropriate Data Collection Instrument and Checking for the Calibration, Validity, and Reliability of Data Collection Instrument Before Collecting the Data During Ph.D. Program in India

H. R. Ganesha [1] & Aithal P. S. [2]

[1] Research Professor, Institute of Management & Commerce, Srinivas University, Mangaluru, India, and Board Member, Gramss Retail Trading Private Limited, Bengaluru, India,
OrcidID: 0000-0002-5878-8844; E-mail: hrganesha@yahoo.co.in

[2] Professor & Vice-Chancellor, Srinivas University, Mangaluru, India,
OrcidID: 0000-0002-4691-8736; E-mail: psaithal@gmail.com

ABSTRACT

Purpose: *The purpose of this article is to explain various types of data collection instruments, and available techniques for checking the calibration, validity, and reliability of the data collection instrument, in turn, guiding Ph.D. scholars to use a well-calibrated, valid, and reliable data collection instrument to ensure high-quality of research output.*

Design/Methodology/Approach: *Postmodernism philosophical paradigm; Inductive research approach; Observation data collection method; Longitudinal data collection time frame; Qualitative data analysis.*

Findings/Result: *As long as the Ph.D. scholars can understand the existing research data collection instruments and realize the need for the development of a new instrument and make mindful choices of instruments in addition to checking their calibration, validity, and reliability before collecting the research data to answer their research question they will be able to determine (on their own) data analysis techniques and appropriate statistical techniques in subsequent stages of the doctoral-level research process to comfortably claim their research findings.*

Originality/Value: *There is a vast literature about research data collection instruments and procedures to check their calibration, validity, and reliability. However, only a few have explained them together comprehensively which is conceivable to Ph.D. scholars. In this article, we have attempted to briefly explain various types of data collection instruments, and available techniques for checking the calibration, validity, and reliability of the data collection instrument.*

Paper Type: *Conceptual.*

Keywords: Research Methodology; Research Design; Research Process; PhD; Ph.D.; Coursework; Doctoral Research; Data Collection Instrument; Human Instrument; Instrument Adoption; Instrument Adaption; Instrument Development; Calibration; Validity; Reliability; Face Validity; Construct Validity; Content Validity; Criterion Validity; Test-retest Reliability; Parallel Forms Reliability; Inter-rater Reliability; Split-half Reliability; Internal Consistency Reliability; Postmodernism

1. BACKGROUND :

Various research studies have identified factors affecting the Ph.D. success rate across the world. "To name a few a) scholar-supervisor/guide relationship; b) mentorship; c) dissertation process; d) role of the department; e) role of peer qualities; f) transformational learning experience provided; g) level of curiosity and interest in reviewing the existing literature; h) planning and time management skills; i) level of creative thinking and writing skills; j) amount of freedom in the research project; k) level of a supportive environment for Ph.D. scholars' well-being; l) higher-education practices; m) supervisors' research capabilities and gender; n) expectations set by the research environment; o) Ph.D. scholars' expectations; p) support network; q) level of Ph.D. scholars' socialization with the research community;

r) Ph.D. scholars' navigation system; s) different terminologies for various components of doctoral-level research are given by different disciplines creating undue confusion in scholars' minds; t) data collection methods which just play the role of data collection and it is just one of the steps of the doctoral-level research process being portrayed as the research methodology/design; u) scholars' inability to identify their genuine interest in a fact/phenomenon/reality/truth/dependent variable, intensive review of existing literature, locating an important research gap, and finally formulating a research question; v) a lower level of clarity about the most important and indispensable step of the doctoral-level research process i.e., choosing an appropriate research philosophical paradigm that lays stepping stones toward answering the research question in a scientific and scholarly way; w) a lower level of clarity about the most important and indispensable step of the doctoral-level research process i.e., choosing an appropriate research approach/reasoning that paves path for decision concerning data collection and analysis; x) a humongous confusion among Ph.D. scholars in India about the difference between research methodology/design and research data collection methods; y) lower level of clarity and the beginning of the Ph.D. journey without a clear understanding of the essence of research data collection time frames; z) lower level of clarity about the right sample size and appropriate sampling techniques" [1-54].

Furthermore, in reality, a majority of stakeholders in the research education system have a lower level of clarity about the most important and indispensable step of the doctoral-level research process i.e., choosing an appropriate research data collection instrument and essence of calibration, validity, and reliability of the chosen data collection instrument. A majority of them guide the Ph.D. scholars to begin the journey without educating the scholars about the most important aspect/objective/purpose of the data collection instrument, calibration, validity, and reliability of the instrument. They also mandate that scholars use certain data collection instruments that are commonly used in a discipline or the one with which they are comfortable. In addition, there is a humongous confusion about i) the data collection instruments; ii) the difference between Mechanical/Electrical/Electronic instruments and Human instruments; iii) the difference between 'Adopted', 'Adapted', and 'Developed' Human instruments; iv) difference between validity and reliability. This lower level of clarity and the beginning of the Ph.D. journey without a clear understanding of these differences is making it difficult for Ph.D. scholars to complete the journey successfully and most importantly if some scholars complete their Ph.D. journey successfully, their awareness about the reasons for choosing a specific data collection instrument is very low. We believe that if the scholars can begin their Ph.D. journey by allocating a higher level of focus and time toward understanding the right data collection instrument their journey will be with a very lower level of complications. But this reality is knowingly or unknowingly, intentionally, or unintentionally suppressed by a majority of stakeholders in the research education system in India. In other words, this *suppressed reality* has resulted in creating humungous confusion among Ph.D. scholars in India about the essence of data collection instruments and the key purpose of checking the calibration, validity, and reliability of the data collection instrument.

One thing Ph.D. scholars must always remind themselves of throughout their Ph.D. journey is the fact that they will be awarded a Ph.D. degree for doing doctoral-level research. Doing doctoral-level research and generating research outputs such as research articles and a thesis determines the probability of success in getting a Ph.D. degree. The first step of the doctoral-level research process is identifying research gaps and formulating a research question, the second one is choosing an appropriate research philosophical paradigm, the third step is choosing an appropriate research approach/reasoning, the fourth step is choosing the appropriate research data collection method choice, the fifth step is choosing an appropriate data collection time frame, the sixth step is to derive the sample size, the seventh step is to choose samples from the research population, the eighth step is to select a data collection instrument, and the ninth step is checking the calibration, validity, and reliability of the data collection instrument [48]. It is thus inevitable and imperative that Ph.D. scholars understand available data collection instruments and know how to check the calibration, validity, and reliability of the data collection instrument. The doctoral-level research which is the single most important requirement of the Ph.D. program is cognitively demanding and intends to create researchers who can create new knowledge or interpret existing knowledge about reality by using different perspectives, paradigms, and reasoning. Knowledge sharing requires autonomy, good quality time, a stress-free brain for deep thinking, and the freedom to look for more meaningful findings. This is the single most important reason for making doctoral-level research flexible wherein the scientific and scholarly world gives autonomy to Ph.D. scholars to formulate their question and answer it within 3-6 years using an appropriate research approach/reasoning. Nevertheless, only 50% of scholars admitted to Ph.D. in India completed, and that too in ten years whether or not they are aware of the importance of reasoning in doctoral-level research [46].

Appropriate sample size and selection of samples from the research population depends upon i) the type of the research question (descriptive; relational; causal) [49]; ii) the research philosophical paradigm (positivism; interpretivism; critical realism; postmodernism; pragmatism) [50]; iii) the research approach/reasoning (deductive; inductive; abductive) [51]; iv) time available for scholars to collect data [46]; v) data collection method and method choice [52]; vi) resources that are available for scholars to collect data [46]; vii) data collection time frame choice [53]; vii) sample size and sampling technique chosen [54]. Choosing a data collection instrument and checking the calibration, validity, and reliability of the data collection instrument before collecting the research data is one of the most important decisions scholars need to make during their Ph.D. journey. Do note that the data collection instrument tells us *'Using What We will Collect Data'* and the calibration, validity, and reliability of the data collection instrument tells us *'The Quality of Data Collected'* [48].

2. OBJECTIVE :

There is humongous confusion among Ph.D. scholars in India about i) data collection instruments; ii) the difference between Mechanical/Electrical/Electronic instruments and Human instruments; iii) the difference between validity and reliability. Furthermore, choosing an appropriate data collection instrument and checking the calibration, validity, and reliability of the data collection instrument before collecting the research data is very important during doctoral-level research as this determines the quality of research and research output. *Owing to such confusion the key objective of this article is to explain various types of data collection instruments, and available techniques for checking the calibration, validity, and reliability of the data collection instrument, in turn, guiding them to use a well-calibrated, valid, and reliable data collection instrument to ensure high-quality of research output*.

3. CHOOSING AN APPROPRIATE RESEARCH DATA COLLECTION INSTRUMENT :

Once the Ph.D. scholars have chosen 'How' (data collection method), 'When' (data collection time frame), 'From How Many' (sample size), and 'From Whom' (sampling technique) to collect the research data. Now they need to finalize 'Using Which Instrument' they will collect the research data and check for the calibration, validity, and reliability of the data collection instrument chosen that will determine 'The Quality of Data Collected' from the samples scholars have chosen from the research population in the previous step of the doctoral-level research [46] [48]. This step is also one of the easiest steps during the doctoral-level research process as the scholars' task is to only choose one or more data collection instruments from many available and check the calibration, validity, and reliability of the instruments with the help of standard procedures and formulas known as 'Facilitators' [47]. There are two categories of data collection instruments in research such as i) Mechanical/Electrical/Electronic Instruments, and ii) Human Instruments [55-69]. We have listed a few known Mechanical/Electrical/Electronic instruments and Human Instruments here.

3.1. Examples of a Few Mechanical/Electrical/Electronic Instruments :

- Accelerometer - Measures physical acceleration.
- Ammeter - Measures the strength of the electric current.
- Anemometer - Measures the speed and velocity of the wind
- Barometer - Measures the atmospheric pressure.
- Bevameter - Measures the mechanical properties of soil.
- Breathalyzer - Measures breath alcohol content.
- Calorimeter - Measures the quantity of heat.
- Cardiogram - Traces movements of the heart. Recorded on a cardiograph.
- Cathetometer - Measures vertical distances.
- Densimeter - Measures the specific gravity of liquids.

- Disdrometer - Measures the size, speed, and velocity of raindrops.
- Dynamometer - Measures electrical power.
- Eudiometer - Measures volume changes in the chemical reaction between gases.
- Evaporimeter - Measures the rate of evaporation.
- Fuel Gauge - Measures fuel levels.
- Glucometer - Measures blood glucose.
- Graphometer - Measures angles.
- Hydrometer - Measures the specific gravity of liquids.
- Hydrophone - Measures sound underwater.
- Hygrometer - Measures humidity in the air.
- Inclinometer - Measures the angle of a slope.
- Interferometer - Measures wave interference.
- Katharometer - Measures the composition of gases.
- Kymograph - Graphically records the physical movements of cells.
- Lactometer - Determines the purity of milk.
- Lux meter - Measures the intensity of light.
- Manometer - Measures the pressure of gases.
- Mercury Barometer - Measures atmospheric pressure.
- Nephelometer - Measures particles in a liquid.
- Nephoscope - Measures the speed and direction of clouds.
- Oximetry - Measures the oxygen level/oxygen saturation of the blood.
- Planometer - Measures area.
- Psychrometer - Measures humidity.
- Rotameter - Measures the pressure of a liquid/gas in a closed tube.
- Saccharometer - Measures the amount of sugar in a solution.
- Seismograph - Measures the intensity of earthquake shocks.
- Spirometer - Measures the lung capacity.
- Taximeter - Measures distance traveled and displacement.
- Thermometer - Measures temperature.
- Viscometer - Measures the viscosity of liquids.
- Voltmeter - Measures the electric potential difference between two points.
- Watt Meter - Measures electrical power.
- Wind Vane - Measures wind direction.
- Zymometer - Measures fermentation

3.2. Examples of a Few Human Instruments/Scales :

- Seashore Rhythm Test - Neuropsychological Tests
- Speech Sounds Perception Test - Neuropsychological Tests
- Stroop Color and Word Test Neuropsychological Tests
- Wisconsin Card Sorting Test - Neuropsychological Tests
- Rorschach Inkblot Test - Psychological and Emotional Testing.
- Social Anxiety Scale - Psychological and Emotional Testing
- Suicide Probability Scale - Psychological and Emotional Testing.
- Trauma Symptom Checklist for Children - Psychological and Emotional Testing.
- Sensory Profile Questionnaire - Behavior Assessment.
- Adaptive Behavior Inventory - Behavior Assessment.
- Vineland Adaptive Rating Scales - Behavior Assessment.
- Battelle Developmental Inventory - Developmental Tests.
- Autism Diagnostic Interview - Developmental Tests.
- Social Communication Questionnaire - Developmental Tests.
- Likert Scale – Measures attitude.

3.3 Instrument Selection:

It is easier to select a Mechanical, Electrical, or Electronic Instrument. However, it is difficult to select a Human Instrument. Scholars are allowed to choose any one of the following ways of finalizing a Human Instrument for research data collection as detailed below.

3.3.1. Instrument Adoption :

In situations when there is an existing questionnaire (instrument) that is appropriate to measure variables of the research question then taking all the questions/items/inventory from an existing questionnaire is allowed and this is known as Instrument Adoption. Scholars are not allowed to change any questions/items, and this is feasible only when the context/environment of the research study/population is the same.

3.3.2. Instrument Adaption :

Taking most of the questions from an existing/proven/reliable questionnaire is known as Instrument Adaptation. Scholars are allowed to make changes to a few existing questions to make them appropriate and match the context/environment of their research study/population.

3.3.3. Instrument Development :

In situations wherein, scholars are unable to Adopt or Adapt an existing questionnaire then creating/developing a new questionnaire with all the questions in it being new is the right way and this is known as Instrument Development. Using a questionnaire developed by scholars makes the questionnaire appropriate and matches the context/environment of the research study/population. In case scholars are in this situation, we suggest they follow the below steps for developing a new instrument. Do note that the only chance for research scholars of Social Sciences, Economics, Management, and disciplines other than Basic Sciences, Engineering, and Technology to own a Patent is to develop a new questionnaire/instrument themselves that is capable of measuring one or more directly unmeasurable variables (also known as Latent variables).

- **Step 1:** Clearly define the dependent variable under study.
- **Step 2**: Identify independent variables we intend to study and frame each question of the instrument to ensure they shall measure the response.
- **Step 3:** Define the population and determine the sample size statistically.
- **Step 4:** Design a questionnaire and while designing questions ensure;
  - Words used in the questions must be in line with the education and maturity level of respondents.
  - Questions and answer options must be clear.
  - Avoid using words that could lead to alternate and misinterpretations of answers.
  - Avoid unethical questions; not feasible; personal; objectionable; or biased words and contexts.
  - Avoid lengthy questions; undefined short forms; difficult words.
  - The number of questions must be balanced between positive and negative responses.
- **Step 5:** Do a Delphi (expert opinion) or full-fledge FGD to get a consensus on the quality of questions.
- **Step 6:** First pilot testing of the instrument.
- **Step 7:** Initial test of the instrument based on pilot responses for validity and reliability.
- **Step 8:** Add, delete, modify, or improvise the questionnaire based on the initial pilot test.
- **Step 9:** Second pilot testing of the questionnaire.
- **Step 10:** A final test of the questionnaire based on the second pilot's responses for validity and reliability.
- **Step 11:** Add, delete, modify, or improvise the questionnaire based on the second pilot.
- **Step 12:** Final distribution of questionnaires to respondents and collection of responses.

Furthermore, if the scholars choose an Electrical/Electronic/Mechanical data collection instrument it is mandatory to check the calibration and reliability. And irrespective of whether a Human instrument is Adopted, Adapted, or Developed, scholars must ensure the Human instrument is checked for reliability and validity with the help of standard procedures and statistical techniques before using the Human instrument to collect responses from respondents/participants/subjects/units of analysis/samples. We strongly recommend scholars belonging to disciplines *other than* Basic/Natural Sciences, Engineering, and Technology attempt to develop a new Human instrument during their Ph.D. program and obtain a Patent as this is the only opportunity for them to get a Patent. Generating an intellectual property right in the form of a patent indicates that a scholar has conducted and delivered high-quality research.

4. VALIDITY AND RELIABILITY OF THE DATA COLLECTION INSTRUMENT :

Once scholars have chosen a data collection instrument, now before collecting the data they are required (mandatory) to check the calibration, validity, and reliability of the data collection instrument.

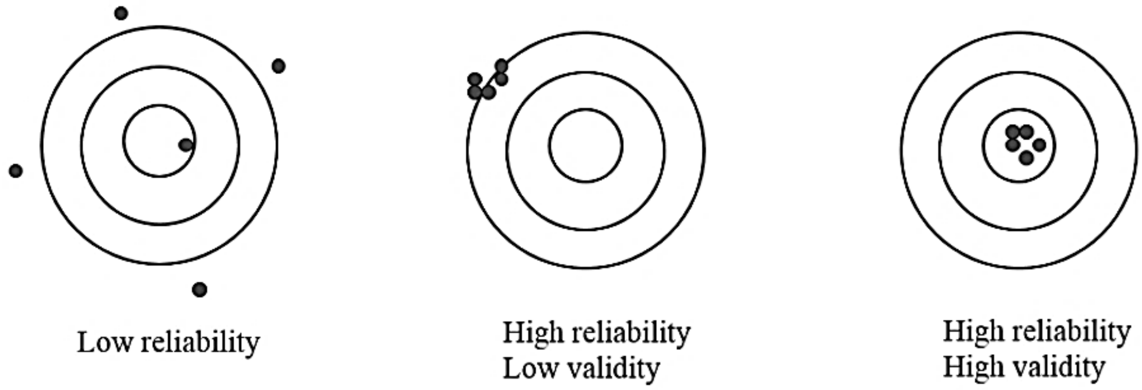

Fig. 1: Difference between validity and reliability [70]

There is a lot of confusion among research scholars about the difference between validity and reliability. An analogy that demonstrates this distinction is shown in figure 1**.** It is comparable to shooting at a target to determine a shooter's proficiency in a particular set of information or skills. Each shot corresponds to one test administration. Reliability and validity are concepts used to evaluate the quality of research. They indicate how well a method, technique, or test (Instrument) measures something. Reliability is about the consistency of a measure, and validity is about the accuracy of a measure. Table 1 explains a few key differences between validity and reliability.

In addition to checking the validity and reliability of the data collection instrument, scholars need also check for calibration in case they have chosen a Mechanical, Electrical, or Electronic Instrument. Scholars must be aware that the validity and reliability of the data collected by them from respondents/participants/subjects/units of analysis/samples are determined by i) type of data collection instrument; ii) calibration, validity, and reliability of the data collection instrument. Scholars are recommended to ensure this before collecting the data. Because it is difficult to recollect the research data in case the data collected are not making any research sense.

Table 1: Difference between validity and reliability [71]

Validity	Reliability
The extent to which the results really measure what they are supposed to measure.	The extent to which the results can be reproduced when the research is repeated under the same conditions.
By checking how well the results correspond to established theories and other measures of the same concept.	By checking the consistency of results across time, across different observers, and across parts of the test itself.
A valid measurement is generally reliable: if a test produces accurate results, they should be reproducible.	A reliable measurement is not always valid: the results might be reproducible, but they are not necessarily correct.

5. CHECKING CALIBRATION OF THE DATA COLLECTION INSTRUMENT :

Calibration defines the accuracy and quality of measurements recorded using a piece of Mechanical, Electrical, or Electronic Instrument. Over time there is a tendency for results and accuracy to drift, particularly when using particular technologies or measuring particular parameters such as temperature and humidity. To be confident in the results being measured there is an ongoing need to service and maintain the calibration of Mechanical, Electrical, or Electronic instruments throughout their lifetime for reliable, accurate, and repeatable measurements. The goal of Calibration is to minimize any measurement uncertainty by ensuring the accuracy of test Mechanical, Electrical, or Electronic instruments. Calibration quantifies and controls errors or uncertainties within measurement processes to an acceptable level. Typically, the accuracy of the standard should be ten times the accuracy of the measuring device being tested. However, an accuracy ratio of 4:1 is acceptable by most standards organizations. A measuring device should be calibrated i) according to the recommendation of the manufacturer of the device, ii) after any mechanical or electrical shock, and iii) periodically (annually, quarterly, monthly) [72-79].

6. CHECKING THE VALIDITY OF THE DATA COLLECTION INSTRUMENT :

There are four steps to check the validity of a Human Instrument (Questionnaire/Test) as listed below [80-90]. Scholars need to be aware that all these steps are mandatory in case they have chosen a Human Instrument that is either 'Adopted', 'Adapted', or 'Developed'.

6.1. Validity Check Step 1 - Face Validity :

Face validity refers to the extent to which an instrument appears to measure what it is intended to measure. A test in which most people (experts) would agree that the Human Instrument (Questionnaire) items (Questions) appear to measure what the instrument is intended to measure would have strong face validity. Here scholars might just get their Human Instrument evaluated by a few experts in the field of their research (Research Supervisor/Guide/Faculty Members/Industry Experts etc).

6.2. Validity Check Step 2 - Construct Validity :

It is the adherence of a measure to existing theory and knowledge of the concept being measured and this check is required when you are Developing a new Human Instrument. Here, scholars will get responses to their new Human Instrument from their pilot samples and also get the responses from their pilot samples using some other existing Human Instruments that are very similar to what scholars are trying to measure. After the collection of responses to both the Human Instruments, scholars shall check the correlation between these two Human Instruments which in turn gives them an indication of the Construct Validity of their new Human Instrument.

6.3. Validity Check Step 3 - Content Validity :

It is the extent to which the measurement covers all aspects of the concept being measured. For example, two constructs 'Gratitude' and 'Forgiveness' must ensure Human Instrument scholars have chosen all the questions/items that are relevant, and none are missed as per the existing literature.

6.4. Validity Check Step 4 - Criterion Validity :

It is the extent to which the result of a measure corresponds to other valid measures of the same concept. For example, if scholars have developed a Human Instrument to understand the attitude of the Ph.D. Scholars toward a Research Methodology Programme, then they can either check 'Regularity of Attendance' (Concurrent Validity) or wait till the final Research Methodology 'Exam Scores' are obtained (Predictive Validity) to check the Criterion Validity.

7. CHECKING THE RELIABILITY OF THE DATA COLLECTION INSTRUMENT :

There are five steps to check the reliability of the data collection instrument which are listed below. scholars need to be aware that all these steps are mandatory irrespective of the type of Human data collection instrument they have chosen, and the Test-retest is mandatory for Electrical/Electronic/ Mechanical data collection Instruments [91-116].

7.1. Reliability Check Step 1 - Test-retest :

The consistency of a measure across time is checked using Test-retest. Here, we intend to check whether we get the same results when we repeat the measurement using the same data collection Instrument. Scholars can calculate the Test-retest reliability using Pearson's correlation formula (1) as shown below. Finding a correlation coefficient for the two sets (Test and Retest) of data is one of the most common ways to find a correlation between the two tests. Test-retest reliability coefficients (also called coefficients of stability/r) vary between 0 and 1, where: 1 is perfect reliability; ≥ 0.9 is excellent reliability; ≥ 0.8 < 0.9 is good reliability; ≥ 0.7 < 0.8 is acceptable reliability; ≥ 0.6 < 0.7 is questionable reliability ≥ 0.5 < 0.6 is poor reliability; < 0.5 is unacceptable reliability; 0: means no reliability.

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (1)$$

r = correlation coefficient

x_i = values of the x-variable in a sample

$\bar{x}$ = mean of the values of the x-variable

y_i = values of the y-variable in a sample

$\bar{y}$ = mean of the values of the y-variable

7.2. Reliability Check Step 2 - Parallel Forms :

The key to this method is the development of alternate test forms that are equivalent in terms of content, response processes, and statistical characteristics. Finding a correlation coefficient for the two sets (Form 1 and Form 2) of data is one of the most common ways to find a correlation between the two Forms. Scholars can calculate the Parallel-forms reliability using Pearson's correlation formula (1). Parallel-forms reliability coefficients (also called coefficients of stability/'r') vary between 0 and 1, where: 1 is perfect reliability; ≥ 0.9 is excellent reliability; ≥ 0.8 < 0.9 is good reliability; ≥ 0.7 < 0.8 is acceptable reliability; ≥ 0.6 < 0.7 is questionable reliability ≥ 0.5 < 0.6 is poor reliability; < 0.5 is unacceptable reliability; 0: means no reliability.

7.3. Reliability Check Step 3 - Inter-rater :

Inter-rater is the consistency of a measure across raters or observers. We intend to check whether we get the same results when different people conduct the same measurement. The more difficult and rigorous way to measure inter-rater reliability is to use Cohen's Kappa, which calculates the percentage of items that the raters agree on while accounting for the fact that the raters may happen to agree on some items purely by chance. Cohen's Kappa formula (2) is shown below.

$$k = (p_o - p_e) / (1 - p_e) \quad (2)$$

p_o: Relative observed agreement among raters

p_e: Hypothetical probability of chance agreement

7.4. Reliability Check Step 4 - Split-half :

This method treats the two halves of a measure as alternate forms. For example, administering a test to a group of individuals; splitting the test in half; correlating scores on one half of the test with scores on the other half of the test. Scholars can use Flanagan's correlation coefficient formula (3) for Split-half Reliability.

$$r = 2\left(1 - \frac{S_a^2 + S_b^2}{S^2}\right)$$

$S_a{}^2$ is the variance of part a
$S_b{}^2$ is the variance of part b
S^2 is the variance of total scores

(3)

7.5. Reliability Check Step 5 - Internal Consistency :

It is the consistency of the measurement itself. Do we get the same results from different parts of a test that are designed to measure the same thing? Scholars can use Cronbach's alpha formula (4) to check internal consistency. Cronbach alpha of ≥ 0.9 is excellent internal consistency; ≥ 0.8 < 0.9 is good internal consistency; ≥ 0.7 < 0.8 is an acceptable internal consistency; ≥ 0.6 < 0.7 is questionable internal consistency ≥ 0.5 < 0.6 is poor internal consistency; < 0.5 is an unacceptable internal consistency.

$$\alpha = \left(\frac{k}{k-1}\right)\left(1 - \frac{\sum_{i=1}^{k} \sigma_{y_i}^2}{\sigma_x^2}\right)$$

k refers to the number of scale items
$\sigma_{y_i}^2$ refers to the variance associated with item i
σ_x^2 refers to the variance associated with the observed total scores

(4)

8. CONCLUSION :

Ph.D. scholars must know that it is very important to ensure they mention the results of the Calibration, Validity, and Reliability of their data collection instrument in a majority of sections of either a research article or their Ph.D. thesis. For example, in the Literature Review Section, mention what have other researchers done to devise and improve methods that are reliable and valid. In the Methodology Section mention, how did the scholars plan their research to ensure the reliability and validity of the measures used? This includes the chosen sample set and size, sample preparation, external conditions, and measuring techniques. In the Results Section, mention the values of reliability and validity alongside the main results. In the Discussion Section, scholars can talk about how reliable and valid their results were. Were they consistent, and did they reflect true values? If not, why not? And finally, in the Conclusion Section, if reliability and validity were a big problem for research findings, it might be helpful to mention them here.

Among the three types of Human instruments available, 'Adoption' and 'Adaption' of an existing instrument are the most preferred by scholars belonging to disciplines *other than* Basic/Natural Sciences, Engineering, and Technology in India. We understand the Ph.D. program is time-bound and hence using and 'Adopted' or 'Adapted' Human Instrument for research data collection during the Ph.D. program is acceptable. But knowingly or unknowingly, intentionally, or intentionally a significant majority of researchers in India use 'Adopted' or 'Adapted' Human Instruments for research data collection even after the completion of the Ph.D. program. *We must be cognizant of the fact that, despite existing data collection instruments, there are yet many variables that are still directly not measurable. Developing a new data collection instrument to measure an unmeasurable variable would be the most desired research output from Ph.D. scholars and Ph.D. holders in India*. The fear among Indian researchers is that Random/Probability sampling techniques require a lot of time investment, they are complicated, and most importantly the research output in the form of research article publications will slow down drastically. The mere pressure on Ph.D. scholars and Ph.D. holders in India to publish a certain number of research articles which is connected to their performance measurement is also one of the key reasons for this. Ph.D. scholars and Ph.D. holders must be aware that a scholarly description, explanation, or claim about a reality/fact/truth/effect/dependent variable and a piece of complete knowledge about reality is complete only when they are derived from collecting research data using multiple data collection instruments with high validity and reliability.

It is the responsibility of every stakeholder in the research environment and system to ensure that the scholars are made aware of every step involved in carrying out doctoral-level research in addition to the purpose, objective, validity, and reliability of various types of research data collection instruments for them to choose an appropriate one to achieve their key research objective during the Ph.D. journey. Designing robust coursework that is intended to create awareness about the essence of research data collection instruments and the calibration, validity, and reliability of the data collection instruments is an appropriate way of fulfilling this responsibility. As long as the Ph.D. scholars can understand the existing research data collection instruments and realize the need for the development of a new instrument and make mindful choices of instruments in addition to checking their calibration, validity, and reliability before collecting the research data to answer their research question they will be able to determine (on their own) data analysis techniques and appropriate statistical techniques to comfortably claim their research findings.

REFERENCES :

[1] Titus, S. L., & Ballou, J. M. (2013). Faculty members' perceptions of advising versus mentoring: Does the name matter?. *Science and Engineering ethics*, *19*(3), 1267-1281. Google Scholar↗

[2] Ali, A., & Kohun, F. (2006). Dealing with isolation feelings in IS doctoral programs. *International Journal of Doctoral Studies*, *1*(1), 21-33. Google Scholar↗

[3] Ali, A., Kohun, F., & Levy, Y. (2007). Dealing with Social Isolation to Minimize Doctoral Attrition- A Four Stage Framework. *International Journal of Doctoral Studies*, *2*(1), 33-49. Google Scholar↗

[4] Spaulding, L. S., & Rockinson-Szapkiw, A. (2012). Hearing their voices: Factors doctoral candidates attribute to their persistence. *International Journal of Doctoral Studies*, *7*, 199. Google Scholar↗

[5] Golde, C. M., & Dore, T. M. (2001). At cross purposes: What the experiences of today's doctoral students reveal about doctoral education, *ERIC Processing and Reference Facility*, 1-62. Google Scholar↗

[6] Golde, C. M. (2005). The role of the department and discipline in doctoral student attrition: Lessons from four departments. *The Journal of Higher Education*, *76*(6), 669-700. Google Scholar↗

[7] Golde, C. M., & Walker, G. E. (Eds.). (2006). *Envisioning the future of doctoral education: Preparing stewards of the discipline-Carnegie essays on the doctorate* (Vol. 3). John Wiley & Sons. Google Scholar↗

[8] Gardner, S. K. (2009). Student and faculty attributions of attrition in high and low-completing doctoral programs in the United States. *Higher education*, *58*(1), 97-112. Google Scholar↗

[9] Gardner, S. K. (2010). Faculty perspectives on doctoral student socialization in five disciplines. *International Journal of Doctoral Studies*, *5*, 39. Google Scholar↗

[10] Solmon, M. A. (2009). How do doctoral candidates learn to be researchers? Developing research training programs in kinesiology departments. *Quest*, *61*(1), 74-83. Google Scholar↗

[11] Nogueira-Martins, L. A., Fagnani Neto, R., Macedo, P. C. M., Citero, V. D. A., & Mari, J. D. J. (2004). The mental health of graduate students at the Federal University of São Paulo: a preliminary report. *Brazilian Journal of Medical and Biological Research*, *37*, 1519-1524. Google Scholar↗

[12] Knox, S., Schlosser, L. Z., Pruitt, N. T., & Hill, C. E. (2006). A qualitative examination of graduate advising relationships: The advisor perspective. *The Counseling Psychologist*, *34*(4), 489-518.

Google Scholar↗

[13] Grady, R. K., La Touche, R., Oslawski-Lopez, J., Powers, A., & Simacek, K. (2014). Betwixt and between: The social position and stress experiences of graduate students. *Teaching Sociology*, *42*(1), 5-16. Google Scholar↗

[14] Russell, J., Gaudreault, K. L., & Richards, K. A. R. (2016). Doctoral student socialization: Educating stewards of the physical education profession. *Quest*, *68*(4), 439-456. Google Scholar↗

[15] Russell, J. A. (2015). Rolling with the punches: Examining the socialization experiences of kinesiology doctoral students. *Research quarterly for exercise and sport*, *86*(2), 140-151. Google Scholar↗

[16] Harding-DeKam, J. L., Hamilton, B., & Loyd, S. (2012). The hidden curriculum of doctoral advising. *NACADA Journal*, *32*(2), 5-16. Google Scholar↗

[17] Mansson, D. H., & Myers, S. A. (2012). Using mentoring enactment theory to explore the doctoral student–advisor mentoring relationship. *Communication Education*, *61*(4), 309-334. Google Scholar↗

[18] Robinson, E. M., & Tagher, C. G. (2017). The companion dissertation: Enriching the doctoral experience. *Journal of Nursing Education*, *56*(9), 564-566. Google Scholar↗

[19] Haynes, K. N. (2008). Reasons for doctoral attrition. *Health*, *8*, 17-4. Google Scholar↗

[20] Mazerolle, S. M., Bowman, T. G., & Klossner, J. C. (2015). An analysis of doctoral students' perceptions of mentorship during their doctoral studies. *Athletic Training Education Journal*, *10*(3), 227-235. Google Scholar↗

[21] Holsinger Jr, J. W. (2008). Situational leadership applied to the dissertation process. *Anatomical Sciences Education*, *1*(5), 194-198. Google Scholar↗

[22] McNamara, J. F., Lara-Alecio, R., Hoyle, J., & Irby, B. J. (2010). Doctoral program issues: Commentary on companion dissertations. *A Doctoral Issues Presentation at the National Council of Professors of Educational Administration* Lexington, KY, August 2, 2006. Google Scholar↗

[23] Carter-Veale, W. Y., Tull, R. G., Rutledge, J. C., & Joseph, L. N. (2016). The dissertation house model: Doctoral student experiences coping and writing in a shared knowledge community. *CBE—Life Sciences Education*, *15*(3), ar34. Google Scholar↗

[24] Devos, C., Boudrenghien, G., Van der Linden, N., Azzi, A., Frenay, M., Galand, B., & Klein, O. (2017). Doctoral students' experiences leading to completion or attrition: A matter of sense, progress and distress. *European journal of psychology of education*, *32*(1), 61-77. Google Scholar↗

[25] Beatty, S. E. (2001). The doctoral supervisor-student relationship: some American advice for success. *The Marketing Review*, *2*(2), 205-217. Google Scholar↗

[26] Carpenter, S., Makhadmeh, N., & Thornton, L. J. (2015). Mentorship on the doctoral level: An examination of communication faculty mentors' traits and functions. *Communication Education*, *64*(3), 366-384. Google Scholar↗

[27] Most, D. E. (2008). Patterns of doctoral student degree completion: A longitudinal analysis. *Journal of College Student Retention: Research, Theory & Practice*, *10*(2), 171-190. Google Scholar↗

[28] Stock, W. A., Siegfried, J. J., & Finegan, T. A. (2011). Completion rates and time-to-degree in economics PhD programs (with comments by David Colander, N. Gregory Mankiw, Melissa P. McInerney, James M. Poterba). *American Economic Review*, *101*(3), 176-88. Google Scholar↗

[29] Wamala, R., Ocaya, B., & Oonyu, J. C. (2012). Extended Candidature and Non-Completion of a Ph. D. at Makerere University, Uganda. *Contemporary Issues in Education Research*, *5*(3), 175-184. Google Scholar↗

[30] https://academy.pubs.asha.org/2011/12/higher-education-practices-that-promote-phd-completion/. Retrieved in October 2022.

[31] Preston, J. P., Ogenchuk, M. J., & Nsiah, J. K. (2014). Peer mentorship and transformational learning: PhD student experiences. *Canadian Journal of Higher Education*, *44*(1), 52-68. Google Scholar↗

[32] Devine, K., & Hunter, K. H. (2017). PhD student emotional exhaustion: the role of supportive supervision and self-presentation behaviours. *Innovations in Education and Teaching International*, *54*(4), 335-344. Google Scholar↗

[33] Van Rooij, E., Fokkens-Bruinsma, M., & Jansen, E. (2021). Factors that influence PhD candidates' success: the importance of PhD project characteristics. *Studies in Continuing Education*, *43*(1), 48-67. Google Scholar↗

[34] Chenevix-Trench, G. (2006). What makes a good PhD student?. *Nature*, *441*(7090), 252-252. Google Scholar↗

[35] Dericks, G., Thompson, E., Roberts, M., & Phua, F. (2019). Determinants of PhD student satisfaction: the roles of supervisor, department, and peer qualities. *Assessment & evaluation in higher education volume 44*(7), 1053-1068. Google Scholar↗

[36] Corsini, A., Pezzoni, M., & Visentin, F. (2022). What makes a productive Ph. D. student?. *Research Policy 51*(10), 104561. Google Scholar↗

[37] Lindvig, K. (2018). The implied PhD student of interdisciplinary research projects within monodisciplinary structures. *Higher Education Research & Development*, *37*(6), 1171-1185. Google Scholar↗

[38] Holbrook, A., Shaw, K., Scevak, J., Bourke, S., Cantwell, R., & Budd, J. (2014). PhD candidate expectations: Exploring mismatch with experience. *International Journal of Doctoral Studies*, *9*, 329. Google Scholar↗

[39] Björkman, B. (2015). PhD supervisor-PhD student interactions in an English-medium Higher Education (HE) setting: Expressing disagreement. *European Journal of Applied Linguistics*, *3*(2), 205-229. Google Scholar↗

[40] Dimitrova, R. (2016). Ingredients of good PhD supervision-evidence from a student survey at Stockholm university. *Utbildning och Lärande/Education and Learning*, *10*(1), 40-52. Google Scholar↗

[41] Sullivan-Bolyai, S., & L'Esperance, S. (2022). Reflections on virtual research conferences and PhD student socialization: The missing link of in-person human connectedness. *Applied Nursing Research*, *64* (April 2022), 151553. Google Scholar↗

[42] Alpert, F., & Eyssell, T. H. (1995). Getting the most from your doctoral program: Advice for the Ph. D. student in finance. *Journal of Financial Education*, 12-20. Google Scholar↗

[43] Groen, J. (2020). *Perceptions of Transformation and Quality in Higher Education: A Case Study of PhD Student Experiences* (Doctoral dissertation, University of Ottawa). Google Scholar↗

[44] Helfer, F., & Drew, S. (2013). A small-scale investigation into Engineering PhD student satisfaction with supervision in an Australian university campus. In *24th Annual Conference of the Australasian Association for Engineering Education* (pp. 1-9). Google Scholar↗

[45] Cunningham-Williams, R. M., Wideman, E., & Fields, L. (2019). Ph. D. Student Development: A Conceptual Model for Research-Intensive Social Work PhD Programs. *Journal of Evidence-Based Social Work*, *16*(3), 278-293. Google Scholar↗

[46] Ganesha, H. R. & Aithal, P. S. (2022). *Doing Ph.D. in India. A Step-by-Step Guide*. First Edition. Notion Press. India & Singapore. ISBN: 9798887832005. Google Scholar↗

[47] Ganesha, H. R. & Aithal, P. S. (2022). The '8Fs' Concept for Simplifying the Complications of Ph.D. Journey in India. *International Journal of Case Studies in Business, IT, and Education*

(IJCSBE), *6*(2), 320-339. Google Scholar↗

[48] Ganesha, H. R. & Aithal, P. S. (2022). The DDLR Model of Research Process for Designing Robust and Realizable Research Methodology During Ph.D. Program in India. *International Journal of Management, Technology, and Social Sciences (IJMTS)*, *7*(2), 400-417. Google Scholar↗

[49] Ganesha, H. R. & Aithal, P. S. (2022). PHDRQ Model for Identifying Research Gaps and Formulating A Research Question During Ph.D. Program in India. *International Journal of Case Studies in Business, IT, and Education (IJCSBE)*, *6*(2). 408-421. Google Scholar↗

[50] Ganesha, H. R. & Aithal, P. S. (2022). Why is it Called Doctor of Philosophy and Why Choosing Appropriate Research Philosophical Paradigm is Indispensable During Ph.D. Program in India?. *International Journal of Philosophy and Languages (IJPL)*, *1*(1). 42-58. Google Scholar↗

[51] Ganesha, H. R. & Aithal, P. S. (2022). Approaching Research in Different Ways. How to Choose an Appropriate Research Approach/Reasoning During Ph.D. Program in India?. International Journal of Philosophy and Languages (IJPL), 1(1). 59-74. Google Scholar↗

[52] Ganesha, H. R. & Aithal, P. S. (2022). How to Choose an Appropriate Research Data Collection Method and Method Choice Among Various Research Data Collection Methods and Method Choices During Ph.D. Program in India?. *International Journal of Management, Technology, and Social Sciences (IJMTS)*, *7*(2), 455-489. Google Scholar↗

[53] Ganesha, H. R. & Aithal, P. S. (2022). When to Collect Data? Choosing an Appropriate Time Frame for Data Collection During Ph.D. Program in India?. *International Journal of Applied Engineering and Management Letters (IJAEML)*, *6*(2), 271-287. Google Scholar↗

[54] Ganesha, H. R. & Aithal, P. S. (2022). Deriving Right Sample Size and Choosing an Appropriate Sampling Technique to Select Samples from the Research Population During Ph.D. Program in India. International Journal of Applied Engineering and Management Letters (IJAEML), 6(2), 288-306. Google Scholar↗

[55] Kielhofner, G., & Coster, W. J. (2017). Developing and evaluating quantitative data collection instruments. *Kielhofner's research in occupational therapy: Methods of inquiry for enhancing practice*, 274-295. Google Scholar↗

[56] Cappa, C., Petrowski, N., & Njelesani, J. (2015). Navigating the landscape of child disability measurement: A review of available data collection instruments. *Alter*, *9*(4), 317-330. Google Scholar↗

[57] Pan, Y., & de La Puente, M. (2005). Census Bureau guideline for the translation of data collection instruments and supporting materials: Documentation on how the guideline was developed. *Survey Methodology*, *6*. Google Scholar↗

[58] Begley, C. M. (1996). Triangulation of communication skills in qualitative research instruments. *Journal of advanced nursing*, *24*(4), 688-693. Google Scholar↗

[59] Thomas, D. B., Oenning, N. S. X., & Goulart, B. N. G. D. (2018). Essential aspects in the design of data collection instruments in primary health research. *Revista Cefac*, *20*, 657-664. Google Scholar↗

[60] Minnaar, L., & Heystek, J. (2013). Online surveys as data collection instruments in education research: A feasible option?. *South African Journal of Higher Education*, *27*(1), 162-183. Google Scholar↗

[61] Westmoreland, H., Bouffard, S., O'Carroll, K., & Rosenberg, H. (2009). Data Collection Instruments for Evaluating Family Involvement. *Harvard Family Research Project*. Google Scholar↗

[62] Haseski, H. I., & Ilic, U. (2019). An investigation of the data collection instruments developed to measure computational thinking. *Informatics in Education*, *18*(2), 297-319. Google Scholar↗

[63] Willimack, D. K. (2013). Methods for the development, testing, and evaluation of data collection instruments. *Designing and conducting business surveys*, 253-301. Google Scholar↗

[64] Johnston, B. D. (2014). Sharing data collection instruments. *Injury prevention*, *20*(2), 73-73. Google Scholar↗

[65] Moyo, T. (2017). Data collection instruments in research: An ontological and epistemological reflection. *Journal of Public Administration*, *52*(Special Issue 1), 285-295. Google Scholar↗

[66] HĂŞMĂŞANU, M. G., BOLBOACĂ, S. D., Jäntschi, L., Zaharie, G. C., & Drugan, T. C. (2014). Design and implementation of data collection instruments for neonatology research. *Applied Medical Informatics*, *35*(4), 35-44. Google Scholar↗

[67] Stecher, C. C., Bricka, S., & Goldenberg, L. (1996). Travel behavior survey data collection instruments. In *Conference Proceedings 10: Conference on Household Travel Surveys: New Concepts and Research Needs* (pp. 154-169). Google Scholar↗

[68] Johannes, C. B., Woods, J., Crawford, S., Cochran, H., Tran, D., & Schuth, B. (2000). Electronic versus paper instruments for daily data collection. *Annals of epidemiology*, *10*(7), 457. Google Scholar↗

[69] Canals, L. (2017). *Instruments for Gathering Data*. Research-publishing. net. La Grange des Noyes, 25110 Voillans, France. Google Scholar↗

[70] Livingston, S. A., Carlson, J., Bridgeman, B., Golub-Smith, M., & Stone, E. (2018). Test reliability-basic concepts. *Research Memorandum No. RM-18-01). Princeton, NJ: Educational Testing Service*. Google Scholar↗

[71] Middleton, Fiona. (2019). https://www.scribbr.com/methodology/reliability-vs-validity/. Retrieved in October 2022.

[72] Sauter, G. (1995). Goniophotometry: new calibration method and instrument design. *Metrologia*, *32*(6), 685. Google Scholar↗

[73] Imaoka, K., Kachi, M., Kasahara, M., Ito, N., Nakagawa, K., & Oki, T. (2010). Instrument performance and calibration of AMSR-E and AMSR2. *International archives of the photogrammetry, remote sensing and spatial information science*, *38*(8), 13-18. Google Scholar↗

[74] Schneider, T., Petersen, O. H., Kildesø, J., Kloch, N. P., & Løbner, T. (1996). Design and calibration of a simple instrument for measuring dust on surfaces in the indoor environment. *Indoor Air*, *6*(3), 204-210. Google Scholar↗

[75] Maletras, F. X., Augousti, A. T., & Mason, J. (2001, November). Construction and calibration of a new design of Fiber Optic Respiratory Plethysmograph (FORP). In *Optomechanical Design and Engineering 2001* (Vol. 4444, pp. 285-293). SPIE. Google Scholar↗

[76] Zhang, L. L., Zhong, C. B., Li, J. G., Niu, H. Y., Ying, Y. L., & Long, Y. T. (2022). A two-step calibration method for evaluation high bandwidth electrochemical instrument. *Journal of Electroanalytical Chemistry*, *915*, 116266. Google Scholar↗

[77] Németh, G., & Pekker, Á. (2020). New design and calibration method for a tunable single-grating spatial heterodyne spectrometer. *Optics Express*, *28*(15), 22720-22731. Google Scholar↗

[78] Zwinkels, J. C. (1996). Colour-measuring instruments and their calibration. *Displays*, *16*(4), 163-171. Google Scholar↗

[79] Giusca, C. L., & Leach, R. K. (2013). Calibration of the scales of areal surface topography measuring instruments: part 3. Resolution. *Measurement Science and Technology*, *24*(10), 105010. Google Scholar↗

[80] Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2004). The concept of validity. *Psychological review*, *111*(4), 1061. Google Scholar↗

[81] Messick, S. (1987). Validity. *ETS research report series*, *1987*(2), i-208. Google Scholar↗

[82] Angoff, W. H. (2013). Validity: An evolving concept. In *Test validity* (pp. 19-32). Routledge. Google Scholar↗

[83] Brewer, M. B., & Crano, W. D. (2000). Research design and issues of validity. *Handbook of research methods in social and personality psychology*, 3-16. Google Scholar↗

[84] Jenkins, J. G. (1946). Validity for what?. *Journal of Consulting Psychology*, *10*(2), 93. Google Scholar↗

[85] Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into practice*, *39*(3), 124-130. Google Scholar↗

[86] Messick, S. (1980). Test validity and the ethics of assessment. *American psychologist*, *35*(11), 1012. Google Scholar↗

[87] Kitagawa, T. (2015). A test for instrument validity. *Econometrica*, *83*(5), 2043-2063. Google Scholar↗

[88] Burton, L. J., & Mazerolle, S. M. (2011). Survey instrument validity part I: Principles of survey instrument development and validation in athletic training education research. *Athletic Training Education Journal*, *6*(1), 27-35. Google Scholar↗

[89] Taherdoost, H. (2016). Validity and reliability of the research instrument; how to test the validation of a questionnaire/survey in a research. *How to test the validation of a questionnaire/survey in a research (August 10, 2016)*. Google Scholar↗

[90] Zamanzadeh, V., Rassouli, M., Abbaszadeh, A., Majd, H. A., Nikanfar, A., & Ghahramanian, A. (2014). Details of content validity and objectifying it in instrument development. *Nursing Practice Today*, *1*(3), 163-171. Google Scholar↗

[91] Kimberlin, C. L., & Winterstein, A. G. (2008). Validity and reliability of measurement instruments used in research. *American journal of health-system pharmacy*, *65*(23), 2276-2284. Google Scholar↗

[92] Bajpai, S., & Bajpai, R. (2014). Goodness of measurement: Reliability and validity. *International Journal of Medical Science and Public Health*, *3*(2), 112-115. Google Scholar↗

[93] Bartko, J. J., & Carpenter, W. T. (1976). On the methods and theory of reliability. *Journal of Nervous and Mental Disease*. Google Scholar↗

[94] Roberts, P., & Priest, H. (2006). Reliability and validity in research. *Nursing standard*, *20*(44), 41-46. Google Scholar↗

[95] Tobias, P. A., & Trindade, D. (2011). *Applied reliability*. CRC Press. Google Scholar↗

[96] Bruton, A., Conway, J. H., & Holgate, S. T. (2000). Reliability: what is it, and how is it measured?. *Physiotherapy*, *86*(2), 94-99. Google Scholar↗

[97] Gwet, K. L. (2008). Intrarater reliability. *Wiley encyclopedia of clinical trials*, *4*. Google Scholar↗

[98] Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence-based nursing*, *18*(3), 66-67. Google Scholar↗

[99] Boone, D. C., Azen, S. P., Lin, C. M., Spence, C., Baron, C., & Lee, L. (1978). Reliability of goniometric measurements. *Physical therapy*, *58*(11), 1355-1360. Google Scholar↗

[100] Morrow Jr, J. R., & Jackson, A. W. (1993). How "significant" is your reliability?. *Research quarterly for exercise and sport*, *64*(3), 352-355. Google Scholar↗

[101] Cohen, L., Manion, L., & Morrison, K. (2017). Validity and reliability. In *Research methods in education* (pp. 245-284). Routledge. Google Scholar↗

[102] Hammersley, M. (1987). Some notes on the terms 'validity'and 'reliability'. *British educational research journal*, *13*(1), 73-82. Google Scholar↗

[103] Adams, R. J. (2005). Reliability as a measurement design effect. *Studies in Educational Evaluation*, *31*(2-3), 162-172. Google Scholar↗

[104] Brink, P. J. (1991). Issues of reliability and validity. *Qualitative nursing research: A contemporary dialogue*, *2*(12), 164-186. Google Scholar↗

[105] Elsayed, E. A. (2012). Overview of reliability testing. *IEEE Transactions on Reliability*, *61*(2), 282-291. Google Scholar↗

[106] Crowder, M. J., Kimber, A. C., Smith, R. L., & Sweeting, T. J. (2017). *Statistical analysis of reliability data*. Routledge. Google Scholar↗

[107] Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in quantitative methods for psychology*, *8*(1), 23. Google Scholar↗

[108] Gisev, N., Bell, J. S., & Chen, T. F. (2013). Interrater agreement and interrater reliability: key concepts, approaches, and applications. *Research in Social and Administrative Pharmacy*, *9*(3), 330-338. Google Scholar↗

[109] Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *The Journal of Strength & Conditioning Research*, *19*(1), 231-240. Google Scholar↗

[110] Yen, M., & Lo, L. H. (2002). Examining test-retest reliability: an intra-class correlation approach. *Nursing research*, *51*(1), 59-62. Google Scholar↗

[111] McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A. (2011). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and social psychology review*, *15*(1), 28-50. Google Scholar↗

[112] Sadowski, C. J., & Gulgoz, S. (1992). Internal consistency and test-retest reliability of the Need for Cognition Scale. *Perceptual and Motor Skills*, *74*(2), 610-610. Google Scholar↗

[113] Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, *10*(4), 255-282. Google Scholar↗

[114] Wagner, E. E., Maloney, P., & Wilson, D. G. (1981). Split-half and test-retest hand test reliabilities for pathological samples. *Journal of Clinical Psychology*, *37*(3), 589-592. Google Scholar↗

[115] Margolis, R. B., Dunn, E. J., & Taylor, J. M. (1985). Parallel-form reliability of the Wechsler Memory Scale in a geriatric population with suspected dementia. *The Journal of psychology*, *119*(1), 81-85. Google Scholar↗

[116] Weir, C. J., & Wu, J. (2002). Parallel-form reliability: A case study of the GEPT intermediate spoken performance test. In *The Proceedings of the Fifth Academic Forum on English Language Testing in Asia* (pp. 29-58). Google Scholar↗

****