Review Article

An In-Depth Examination of Validity Assessment: Exploring Diverse Methodologies and Dimensions of Validity in Social Research Studies

Abstract

Validity refers to the extent to which a test accurately measures what it claims to measure. Validity is a matter of degree. Face validity refers to an indicator that seems to measure its construct "on its face". Content validity is a form of judgmental validity, that involves determining if a test adequately represents the content domain. Criterion-related validity methods include concurrent and predictive approaches. Construct validity assesses how well theoretical ideas are translated into measurable variables. Face validity involves examining whether a measurement appears suitable based on its operationalization on the face of it. Lawshe's content validity ratio quantifies expert agreement on the essentiality of individual items in a test or scale, while the content validity index considers overall expert agreement on the representativeness of items collectively within a test or scale. It considers the agreement among experts regarding the content validity of the entire set of items. Cohen's Kappa is used to assess the level of agreement between two or more raters or observers when coding or categorizing data. The kappa coefficient is particularly useful when chance agreement among raters is possible. Item Objective Congruence evaluates the efficacy of items in measuring one or more objectives, accomplished by an unbiased expert panel provides evidence of content validity. The most significant assessment at this stage is determining if items and objectives are congruent. Concurrent and predictive validity are assessed using Pearson's correlation and Spearman rank correlation coefficients respectively. The Multi-trait Multimethod Matrix (MTMM), evaluates the construct validity of measures. Validity is essential for ensuring credible and meaningful research findings by enhancing the trustworthiness of the measuring instruments.

Keywords: Validity,Lawshe, CVR, CVI, Kappa, IOC, Pearson, Spearman rank, MTMM

Introduction

Validity refers to the extent to which a test accurately measures what it claims to measure (Kelley, 1927). Thecommonest definition of validity is epitomized by the question: Are we measuring what we thinkwe are measuring? The emphasis in this question is on what is being measured.For example, ateacher has constructed a test to measure understanding of scientific procedures and has includedonly factual items about scientific procedures in the

test.Thetestisnotvalid,becausewhileitmay be reliablein measuring the students' factual knowledge of scientific procedures, it does not measure theirunderstandingofsuchprocedures.Inotherwords,itmaymeasurewhatitmeasuresquitewell,butitdoes notmeasurewhattheteacherintendedittomeasure.(Kerlinger,1986).“Onevalidates,not a test,butaninterpretationofdata arisingfroma specifiedprocedure” (Cronbach,1971).Onedoesnotassessthevalidityofanindicatorbutrathertheusetowhichitisbeingput.Fore xample,an intelligence test may be valid for assessing the native intellectual potential of students, but itwouldnotnecessarilybevalidforotherpurposes,suchasforecastingtheirlevelofincomeduringadulthood (Nunnally, 1978). Just as reliability is a matter of degree, also is validity. Thus, theobjective of attaining a perfectly valid indicator-one that represents the intended, and only theintended, concept is unachievable. Instead, validity is a matter of degree, not an all-or-noneproperty. Moreover, just because an indicator is quite reliable, this does not mean that it is alsorelatively valid. Reliability is an empirical issue, focusing on the performance of empirical measures. Validity, in contrast, is usually more of a theoreticallyoriented issue because it inevitably raises the question, "valid for what purpose?" (Carmines &Zeller,2007)

Review of Literature

Herche (1992) found that the predictive validity of Shimp and Sharma’s (1987) CETSCALEis tested in a nationwide mail survey. The scale is shown to be a much stronger predictor ofimportbuyingbehavior.

Yaghmaie (2003) concluded that for content validity two judgments are necessary: themeasurableextentofeachitemfordefiningthetraitsandthesetofitemsthatrepresentsallaspectsofthetrait s.ItwasfoundthatoutOf38items,thosewithCVIover0.75remainedandtherestwerediscardedresultingina 25-itemscale.

Dellinger (2005) placed construct validity, or the meaning of measures, as the central focusor core of validity encompassing all forms of validity evidence. Any information provided asvalid evidence through traditional means serves to bolster or negate arguments that data from atool,process,observation,instrument,etc.,arevalidindicatorsoftheconstruct ofinterest.

Watson *et al*. (2008) reported the validity of the Inventory of Depression and AnxietySymptoms(IDAS;Watsonetal.,2007)intwo samples(306collegestudents,and605psychiatricpatients). The IDAS scales showed strong convergent validity about parallel interview-based scores on the Clinician Rating version of the IDAS (IDAS-CR); the mean convergentcorrelationswere.51and .62 inthe studentand patientsamples,respectively.

Handageand Chander (2021)found the content validity is assessed in two

rounds,inthefirstroundtheContentValidityRatio(CVR)iscomputedandtheitemswitha CVR critical value of 0.407 or above were accepted and subjected to the computation of Content ValidityIndex(CVI)inthesecondround.Modifiedkappavalueswheredeterminedtoovercomethechancefactor.TheItem-ContentValidityIndex(I-CVI)valueswerecomputedto finalizetheitemsinthefinal scale while the Scale-Content Validity Index (S-CVI) was determined to check for the stabilityof each dimension as well as the scale as a whole. The CVI value above 0.78 is considered fitforthescale.

Priyadarshani *et al.*(2021) concluded that the Item discrimination index is used for thediscriminant validity,allitemscoulddiscriminatebetween the highandlowgroupssignificantly($p<0.01$), and the final test compromised of a total of 12 items with discriminant validity.

Singh *et al.* (2023) reported that the concurrent validity of the Social Cohesiveness Rating Scale(SCRS)isexaminedbyadministeringSpearmanrank-ordercorrelationbetweenthemeanscoreof the two raters completing each scale with the mean score of the other scale. The correlationbetween SCRS and Group Cohesion Scale (GCS) is 0.75 ($p < 0.001$), and this high significantcorrelationimpliesagoodconsiderationofthevalidityofthe SCRS.

Methodology

The research is descriptive. The study is conceptual and based on secondary data collected from journals, books, websites, and previous studies. The paper focuses on the different methods of assessing the validity of the research instruments/tools like scale, test, etc.

2. Types of validity (Test, Scale, and indexes**)**

i. **Face validity**
ii. **Content validity**
iii. **Criterion Related validity**
iv. **Construct validity**

2.1 Face validity

In face validity, you look at the operationalization and see whether "on its face" seemslike a good translation of the construct. This is probably the weakest way to try to demonstrateconstruct validity. We can improve the quality of face validityassessmentconsiderablybymakingitmoresystematic.Forinstance,ifyouaretryingtoassess theface validity of a math ability measure, it would be more convincing if you sent the test to acarefully selected sample of experts on math ability testing and they all reported back with thejudgment thatyourmeasure appearstobe agoodmeasure ofmathability(Jain,2005).

2.2 Content validity

Content validity also known as logical validity is the representativeness or the samplingadequacy of the measuring instrument. It asks the question "To what extent the content of

thetest/scale is representative of the domain of the psychological object under investigation?" Forexample,whenatrainerisinterested inmeasuringtheeffect oftrainingontheknowledgelevelofthefarmer,hewillinitiallyconstructaknowledgetestandsubjectittocontentvaliditytodetermine to what extent the questions/items in the knowledge test are representative of thetrainingcoursecontent.

2.2.1PreconditionsforContentValidation

1. Eachitemmustbejudgedseparatelyforitspresumedrelevancetothepropertybeingmeasured.
2. JudgementshouldbedonebyexpertswhohaveexperienceinstudyingJobserving/experiencingthephenomenonunderinvestigation.
3. Itis notsufficientifalltheitemsaremeasuringwhattheyweresupposedtomeasure.
4. It should be ensured that the range items are capable of catching every dimension of theconstructunderinvestigation

Content validityisjudgmentalvalidityanditisnota comprehensivemeasureofvalidity.Henceinmostcases,contentvalidityalonecannotbeusedasameasureofvalidity.Itmustbecombinedwithothermethodsofvalidationforbetterresults(Patil&Aditya,20016).

2.3 Criterion-Related Validity

Criterion validity is the idea that a valid test should relate closely toother measures of the same theoretical construct. A valid test of intelligenceshould correlate highly with other intelligence tests. It should alsocorrelate with behaviors that are considered to require intelligence, suchas doing well in school(White and McBurney, 2012). Criterion-related validity is of two types one is concurrent validity and the second is predictive validity. Concurrentvalidityisusedgenerallywhenatestisconstructed toreplace an existing test that measures the same construct. Predictive validity is generally usedto demonstrate that a given measuring instrument makes accurate predictions about the construct it is supposed to measure. In the case of predictive validity, one does not care about what is beingmeasured by instrument but one cares for its predictive ability. It is very concerned about thepracticalproblemandoutcomes.

2.4 ConstructValidity(ConvergentandDiscriminantValidity)

The construct validity approach is more complex than other forms of validity. White and McBurney(2012) defined construct validity as the test should measure whatever theoreticalconstruct it supposedly tests, and not something else. There are several ways to determine whether a test generates data that has construct validity.

1. The test should measure whatever theoretical construct it supposedly tests, and not something else. For example, a test of leadership ability should not test extraversion.
2. A test that has construct validity should measure what it intends to measure but not measure theoretically unrelated constructs. For example, a test of musical aptitude should not require

too much reading ability.

3. A test should prove useful in predicting results related to the theoretical concepts it is measuring. For example, a test of musical ability should predict who will benefit from taking music lessons, should differentiate groups who have chosen music as a career from those who haven't should relate to other tests of musical ability, and so on.

There are two types of construct validity— 'convergent validity' and 'divergent validity' (or discriminant validity). In convergent validity, the degree to which the operationalization is similar to (convergeson) other operationalizations that it theoretically should be similar to.Indiscriminant validity, the degree to which the operationalization is not similar to (diverges from)otheroperationalizationsthatittheoreticallyshouldnot besimilarto(Jain,2005).

3. Methods for Assessing various forms of Validity

3.1Lawshe'sContentValidityRatio

The Content Validity Ratio (CVR) is used to assess the content validity of individual items in a test or scale. It was introduced by Lawshe in 1975. The CVR determines whether each item is essential for measuring the construct being assessed. To calculate the CVR, a panel of experts rates each item as either "essential" or "not essential" for the test.

The CVR formula is as follows: CVR = (ne - (N/2)) / (N/2)

Where:

ne = number of experts who rate the item as "essential"

N = total number of experts

The CVR ranges from -1 to +1, and a positive value indicates that the item is considered essential by at least some experts. The commonly used criterion is that an item is considered essential if the CVR is greater than or equal to 0.62. The CVR (content validity ratio) is a linear transformation of a proportional level of agreement on how many 'experts' within a panel rate an item 'essential'. Itcould readily be seen whether the level of agreement amongst panel members is greater than50%.CVRvaluesrangebetween-1(perfectdisagreement)and+1(perfectagreement)withCVRvalues above zero indicating that over half of the panel members agree on an item essential (Ayre andScally,2014).The final CVR computation is dependent on the number of judges or experts who assessthe items. Any item that is perceived by more than half of experts to be essential has some degreeofcontentvalidity.

Table. 1:MinimumvalueofCVR(Lawshe,1975)

No.ofjudges	Minimumvalue
5	.99
6	.99
7	.99

8	.78
9	.75
10	.62

CVRiscalculatedinthe followingway: $CVR = \frac{ne - \frac{N}{2}}{\frac{N}{2}}$

Where;

CVR =content validity ratio

ne = Number of panel members indicating an item 'essential'

N = Number of panel members

Table. 2: Calculation of content validity ratio (CVR)

Item	Expert1	Expert2	Expert3	Expert4	Expert5	CVR
1	1	1	0	0	1	0.2
2	1	1	1	1	1	1.0
3	1	1	1	0	1	0.6
4	1	1	1	0	1	0.6
5	1	1	1	1	0	0.6

CVRoffirstitem:

$$CVR = \frac{3-5/2}{5/2} \qquad \text{CVR} = 0.2$$

3.2Content Validity Index (CVI)

The CVI is a measure used to assess the relevance and representativeness of items collectively within a test or scale. It considers the agreement among experts regarding the content validity of the entire set of items. The CVI is calculated for all individual items (I-CVI) and the overall scale (S-CVI). For CVI, the panel of experts is asked to rate each scale item in terms of its relevance to the underlying construct. A 4-point scale is used to avoid a neutral point. The four points used along the item rating continuum were 1= not relevant, 2 = somewhat relevant, 3 = quite relevant, and 4 = highly relevant.

I-CVI = (number of experts giving a rating of 3 or 4) / (total number of experts)

The S-CVI ranges from 0 to 1, with higher values indicating greater content validity. A commonly used threshold for an acceptable level of content validity is an S-CVI of 0.80 or higher, meaning that at least 80% of the experts agree on the relevance or representativeness of each item (Shrotryia and Dhanda, 2019).

Table. 3: Computation of content validity index (CVI)

Item	E1	E2	E3	E4	E5	E6	A	I-CVI
1	1	1	1	1	1	1	6	1.00
2	1	1	1	0	1	1	5	0.83
3	1	1	1	1	1	1	6	1.00
4	1	0	1	1	0	1	4	0.83
5	1	1	1	1	1	1	6	1.00

I-CVI: item content validity index,S-CVI: scale content validity index,A: no. in agreement

I-CVI = Rating given by the expert / total number of experts

I-CVI = 6/6 = 1

I-CVI should be 1.00 in case of five or fewer judges and case of six or more judges; I-CVI should not be less than 0.78.

S-CVI = Σ (I-CVI) / n

S-CVI = 4.49 / 5

S-CVI (Average) = 0.898 (accepted).

where n is the total number of items

It is recommended that a minimum S-CVI should be 0.8 for reflecting content validity.

3.2.1KappaStatisticcoefficient/Cohen'skappa(JacobCohen,1968).

The Kappa coefficient, also known as Cohen's kappa, was developed by Jacob Cohen, an American statistician and psychologist. Cohen introduced the kappa statistic as a measure of inter-rater agreement or reliability for categorical variables, it has been widely used in various fields to assess the level of agreement between two or more raters or observers when coding or categorizing data. The kappa coefficient is particularly useful when chance agreement among raters is possible.

The Kappa coefficient ensures a better understanding of content validity as it removes any random chance agreement. Kappa statistic is a consensus index of interrater agreement that supplements CVI to ensure that the agreement among experts is beyond chance. Computation of Kappa Statistic requires the calculation of the probability of chance agreement Pc.

Kappa statistic is then calculated as K = (I-CVI – Pc) / (1 – Pc). Evaluation criteria for Kappa are that values above 0.74, between 0.6 and 0.74, and the ones between 0.4 and 0.59 are considered to be excellent, good, and fair, respectively

Table. 4: Estimation of kappa statistic coefficient

Item	E1	E2	E3	E4	E5	E6	A	I-CVI	Pc	K
1	1	1	1	1	1	1	6	1.00	0.015625	1.00
2	1	1	1	0	1	1	5	0.83	0.09375	0.81

3	1	1	1	1	1	1	6	1.00	0.015625	1.00
4	1	0	1	1	0	1	4	0.66	0.9375	Negative
5	1	1	1	1	1	1	6	1.00	0.015625	1.00

Pc = probability of chance agreement, K= kappa statistic

$Pc = [N! / A! (N – A)!] \times 0.5^N$.

Where;

N = number of experts in the panel,

A = number of experts in the panel who agree that the item is relevant.

$Pc = [(6! / 6!) (6-6)!] * 0.5^6$

$Pc = [(6*5*4*3*2*1 / 6*5*4*3*2*1) (0)!] * 0.5^6$

Pc = (1*1) * 0.015625

Pc = 0.015625

Kappa statistic coefficient for the first statement

K = (I-CVI – Pc) / (1 – Pc)

Where;

I-CVI = Item Content Validity Ratio

Pc = Probability of chance agreement

K = (1- 0.015625) / (1- 0.015625)

K =1

3.3Item Objective Congruence Index

An evaluation of the efficacy of items in measuring one or more objectives, accomplished by an unbiased expert panel provides evidence of content validity. The most significant assessment at this stage is determining if items and objectives are congruent. The remaining item analyses are meaningless if there is insufficient proof that the items are measuring what they are supposed to measure. The index of item-objective congruence (IOC) introduced by Rovinelli& Hambleton (1977) is one method to quantitatively measure content experts' judgments of items to evaluate the fit between test items and the table of specifications. IOC is a process in which SMEs (Subject Matter Experts) rate individual items on the degree to which they agree or do not agree with the specific objectives listed by the test developer (Turner *et al*., 2003). Accordingly, an expert evaluates each item by giving a rating of +1 for clearly measuring the objective, -1 for not measuring, or 0 for the unclear objective. After the experts rate the items, the results are calculated to create the indices of IOC for each item on each objective. The choice of a cutoff score for this index to separate "good" from "bad" items can be based on some absolute standard relating to specific proportions of perfect ratings for the items. For example, if one-half of the content

specialists judged an item to be a perfect match to an objective, while the others were not able to make a decision, the computed value of the index would be .50. Thus, test constructors obtaining I' values of .50 would know that at a minimum, at least 50 percent of the content specialists gave a perfect rating to the item.

In a condition where an item measures more than one objective, the multidimensional item formula simplified by Crocker and Aligna is utilized to evaluate the similarity between an item and a set of objectives. The formula for the multidimensional item given by Crocker and Aligna is as follows:

$$I'_{ik}=(M-1)S_{ik}-S'_{ik}/2N-(M-1)$$

Where;

I_{ik} is the index of item-objective congruence for item i and objective k

M is the number of objectives

N is the number of judges

S_{ik} is the sum of the item ratings assigned to objective k

S'_{ik} is the sum of the item ratings assigned to all objectives, except objective k

To illustrate the calculation of the index of item-objective congruence, consider the following example for Item 1 of a test with five objectives or domains. Four judges have been asked to rate each item relative to each of the domains. Their ratings appear in the Table below.

Table. 5: Item objective congruence for item 1 with 5 objectives

Judge	**Objectives**				
	1	**2**	**3**	**4**	**5**
A	-1	+1	-1	0	-1
B	-1	+1	-1	-1	-1
C	-1	+1	-1	-1	-1
D	-1	+1	-1	+1	-1
S_i	-4	+4	-4	-1	-4

For this example, the item-objective congruence for Item 1 and Objective 2 is calculated. Therefore, $I_{ik} = I_{12}$; M = 5; N = 4; $S_{12} = 4$; and $S'_{12} = -13$.

Substitution of the above values into the formula for the index yields:

$$I_{12}= \frac{\overline{(5-1)\,(4)\,-(-13)}}{2(4)\,(5-1)}$$

I_{12}=0.906

3.4Pearson's Correlation for Concurrent Validity

The Pearson correlation measures the strength of the linear relationship between two variables. It has a value between -1 to 1, with a value of -1 meaning a total negative linear correlation, 0 being no correlation, and + 1 meaning a total positive correlation.

Pearson's correlation

$$\mathbf{r} = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

Where:

r = Pearson correlation coefficient

x = Values in the first set of data

y = Values in the second set of data

n = Total number of values.

Table. 6: Computation of Person's correlation for concurrent validity

Respondents	Newley Constructed test (x)	Standardized Test (y)	x^2	y^2	xy
1.	8	7	64	49	56
2.	6	7	36	49	42
3.	8	9	64	81	72
4.	7	7	49	49	49
5.	9	9	81	81	81
6.	5	6	25	36	30
7.	8	8	64	64	64
8.	6	7	36	49	42
9.	9	9	81	81	81
10.	5	5	25	25	25
	$\sum x = 71$	**$\sum y = 74$**	**$\sum x^2 = 525$**	**$\sum y^2 = 564$**	**$\sum xy = 542$**

Calculating the values

$$\mathbf{r} = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

$$\mathbf{r} = \frac{10(542) - (71)(74)}{\sqrt{[10*525 - (71)^2][10*564 - (74)^2]}}$$

r = 0.89

- The newly constructed test has high concurrent validity.

3.5Spearman's Rank Correlation for Predictive Validity

Spearman's rank correlation measures the strength and direction of association between two ranked variables. It gives the measure of monotonicity of the relation between two variables i.e., how well the relationship between two variables could be represented using a monotonic function.

Spearman's rank coefficient:

$$\boldsymbol{\rho} = 1 - \frac{6 \sum d_i^2}{n\,(n^2 - 1)}$$

Where;

$\boldsymbol{\rho}$ = Spearman rank correlation

d_i = difference between the two ranks of each observation

n = number of observations

Table. 7: Estimation of Spearman rank correlation for predictive validity

Applicants	Rank in Aptitude test	Rank in Job performance	Difference in ranks (d)	d^2
1.	6	5	1	1
2.	3	2	1	1
3.	1	3	-2	4
4.	8	6	2	4
5.	4	7	-3	9
6.	2	1	1	1
7.	10	9	1	1
8.	9	8	1	1
9.	5	4	1	1
10.	7	10	-3	9
				$\sum d_i^2 = 32$

$$\boldsymbol{\rho} = 1 - \frac{6 \sum d_i^2}{n\,(n^2 - 1)}$$

$$\boldsymbol{\rho} = 1 - \frac{6 * 32}{10\,(10^2 - 1)}$$

$$\boldsymbol{\rho} = 0.806$$

a) Predictor = Aptitude test

b) Criterion = Job performance

c) The test has high predictive validity

3.6Multi-traitMulti-method Matrix (MTMM)

The Multi-trait Multimethod Matrix (MTMM) is an approach to assessing the construct validity of a set of measures in a study. It was developed in 1959 by Campbell and Fiske. Along with the MTMM, they also introduced two new types of validity, convergent and discriminant as subcategories of construct validity. We can assess both convergent and discriminant validity using the MTMM. To be able to claim that your measures have construct validity, you have to demonstrate both convergence and discrimination.

To construct an MTMM, you need to arrange the correlation matrix by concepts within methods. The Table shows an MTMM for three concepts (traits A, B, and C) each of which is measured with three different methods (1, 2, and 3) Note that you lay the matrix out in blocks by method. Essentially, the MTMM is just a correlation matrix between your measures.

The MTMM idea provided an operational methodology for assessing construct validity. In one matrix, it was possible to examine both convergent and discriminant validity simultaneously. By its inclusion of methods on an equal footing with traits, Campbell and Fiske stressed the importance of looking for the effects of how we measure in addition to what we measure. And, MTMM provided a rigorous framework for assessing construct validity.

Table. 8: MMTM matrix for three traits and three methods (Schmitt and Stults, 1986; O'Leary-Kelly and Vokurka, 1998)

		$Method_1$			**$Method_2$**			**$Method_3$**		
	Traits	A_1	B_1	C_1	A_2	B_2	C_2	A_3	B_3	C_3
$Method_1$	A_1	**A_1A_1**								
	B_1	A_1A_2	**A_2A_2**							
	C_1	A_1A_3	A_2A_3	**A_3A_3**						
$Method_2$	A_2	**A_1B_1**	A_2B_1	A_3B_1	B_1B_1					
	B_2	A_1B_2	A_2B_2	A_3B_2	B_1B_2	B_2B_2				
	C_2	A_1B_3	A_2B_3	A_3B_3	B_1B_3	B_2B_3	B_3B_3			
$Method_3$	A_3	A_1C_1	A_2C_1	A_3C_1	B_1C_1	B_2C_1	B_3C_1	C_1C_1		
	B_3	A_1C_2	A_2C_2	A_3C_2	B_1C_2	B_2C_2	B_3C_2	C_1C_2	**C_2C_2**	

Mono-trait Mono-method

Mono-trait Hetero-method (Convergent Validity)

Hetero-trait Hetero-method (Divergent validity)

Hetero-trait Mono-method (Method Variance)

	C_3	A_1C_3	A_2C_3	A_3C_3	B_1C_3	B_2C_3	B_3C_3	C_1C3	C_2C_3	$\mathbf{C_3C_3}$

1. For Convergent validity: MTHM correlations should be relatively large and significantly different from zero.
2. For Discriminant validity: It involves three different criteria, which are as follows
   a. First, any MTHM correlation must be significantly larger than any of the other correlations located onthe same row and column.
   b. Second, a variable should correlate higher with the same variable measured by different methods MTHMcorrelations, than with different variables measured by the same method HTMMcorrelations.
   c. Third, the pattern of correlations between HTMM and HTHM correlation triangles must be the same.

- Failure to meet the criteria indicates that the measures are corrupted potentially by method bias.

Conclusion

Validity is crucial for ensuring accurate, credible, and meaningful researchfindings. It instills confidence in measurements and assessments, allowingresearchers to draw accurate conclusions and make informed decisions. Validity ensures thatresearchers measure what they intend to measure, enhancing the relevance of the results. Itenhancesthecredibility of researchby demonstratingthetrustworthinessofthe instruments. Understandingandensuringvalidityisessentialforproducinghigh-qualityvalidresearchoutcomesthatinformdecision-makingand advanceknowledgein variousfields.

References

Aravamudhan, N.R. and Krishnaveni, R. 2015. Establishing and reporting content validity evidence of training and development capacity building scale (TDCBS). *Management: Journal of Contemporary Management Issues*, 20(1): 131-158.

Bodie, G.D., Jones, S.M., Vickery, A.J., Hatcher, L. and Cannava, K. 2014. Examining the construct validity of enacted support: A multitrait–multimethod analysis of three perspectives for judging immediacy and listening behaviours. *Communication Monographs*, 81(4): 495-523.

Campbell, D. T. and Fiske, D. W. 1959. Convergent and discriminant validation by the multi-trait multi-method matrix.*Psychological Bulletin*, 56(2): 81-105.

Carmines,E.G. andZeller,R.A.*ReliabilityandValidityAssess*ment,California: SagePublications,2007,p

72.

Cohen, J. 1968. Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4): 213.

Crocker,L.andAlgina,J.*Introductiontoclassicalandmoderntesttheory*.New York: Wadsworth Publishing Co Inc, 2006, p 527.

Cronbach,L.J.Testvalidation.InThorndike, R.L.(Ed.).*EducationalMeasurement*,2nd edition, USA: AmericanCouncilonEducation, 1971, pp. 443-507.

Cronbach,L.J.andMeehl,P.E.1955.Constructvalidityinpsychologicaltests.*PsychologicalBulletin*,52(4): 281.

Graham, J.C. and Garton, B.L. 2003. Certification measures: Are they predictive of secondary agriculture teacher performance? *Journal of Agricultural Education*, 44(3): 54-65.

Handage, S. and Chander, M. 2021. Development of an instrument for measuring the student learning outcomes: A content validation process. *Indian Journal of Extension Education*, **57**(3): 1-7.

Ibiyemi, A., Mohd A. Y., Daud, M.N., Olanrele, S. and Jogunola, A. 2019. A content validity study of the test of valuers' support for capturing sustainability in the valuation process in Nigeria. *Pacific Rim Property Research Journal*, 25(3): 177-193.

Jain,M.K.*ResearchMethodologyandStatisticalTechniques*,New Delhi:ShreePublishers&Distributors,2005. p 290.

Kelley, T. L. *Interpretation of Educational Measurements*, New York: World Book Company, 1927, p 392.

Kerlinger, F.N.*FoundationofBehavioralResearch*,New York: Holt,RinehartandWinston, 2003, p 741.

Lawshe, C.H. 1975. A quantitative approach to content validity. *Personnel Psychology*, 28(4): 563-575.

Nunnally,J.C.*PsychometricTheory*,New York: McGraw-Hill, 1978, p 752.

O'Leary, S.W. and Vokurka, R.J. 1998. The empirical assessment of construct validity. *Journal of Operations Management*, 16(4): 387-405.

Patil, S. and Aditya. *Research Methodology in Social Sciences*, New Delhi: New India PublishingAgency, 2016, p 148.

Rovinelli, R. J., & Hambleton, R. K. 1977. On the Use of Content Specialists in the AssessmentofCriterion-ReferencedTestItemValidity. *DutchJournalofEducationalResearch*,2:49-60.

Schmitt, N. and Stults, D.M. 1986. Methodology review: Analysis of multitrait-multimethod matrices. *Applied Psychological Measurement*, 10(1): 1-22.

Schwab, S., Zurbriggen, C.L. and Venetz, M. 2020. Agreement among student, parent and teacher ratings of school inclusion: A multitrait-multimethod analysis. *Journal of School Psychology*, 82: 1-16.

Shrotryia, V.K. and Dhanda, U. 2019. Content validity of assessment instrument for employee engagement. *Sage Open*, 9(1): 1-7.

Singh, S.S., Singh, R.J., Devarani, L., Hemochandra, L. and Singh, R. 2023. Development of reliability and validity of social cohesiveness rating scale (SCRS). *Indian Journal of Extension Education*, 59(1): 139-141.

Turner, Ronna, C., & Carlson, L. 2003. Indexes of Item-Objective Congruence for Multidimensional Items. *International Journal of Testing*, 3(2): 163-171.

White, T.L. and McBurney, D.H. 2012. *Research methods*. 9th ed., USA: Cengage Learning, p 481.

Yaghmaie, F. 2003. Content validity and its estimation. *Journal of medical education*, 3(1): 25-27.