Validity evidence based on content: Controversies and quantification

Agung Santoso
Alice Whita Savira
Robertus Landung Eko Prihatmoko

University of Sanata Dharma

Agung Santoso is currently an assistant professor in the Psychology Department at the University of Sanata Dharma. He earned his Ph.D. in Quantitative Psychology from the University of Notre Dame in 2019. His research interests are in quantitative methods - particularly in good practice in statistical analysis - statistical modeling - particularly in measurement models and longitudinal study -, learning and academic analytics, and causal inference. Most of his works are in good practice in statistical analyses, including the ones presented at the recent conference.

Abstract

Validity is a crucial issue in test development because it represents the accuracy of test score interpretation in describing the measured attribute or construct. Among validity evidence that can be collected to support test score interpretation, the one based on test content incited controversies. The content-based validity evidence includes analyses of the relationship between test content and the construct the test purported to measure. Some authors opposed the use of such evidence in the validation process, arguing that the evidence can only provide hypotheses to be tested in the subsequent processes. Other authors argued that the content validation process provided meaningful information regarding test validity, particularly tests based on content such as achievement tests. Unfortunately, literature regarding such issues hitherto has not addressed the controversies adequately. Furthermore, several techniques for quantifying the content-based validity evidence have been proposed without any other article reviewing and evaluating them in comparison with the others. The current study was conducted to examine propositions made by prominent authors regarding validity based on test content and to compare proposed quantification techniques. Using Bielefeld Academic Search Engine, 1,841 articles with "content validity" in their titles, however only 28 of them met inclusion criteria and thus reviewed. The reviews showed that there were three positions regarding content-based validity evidence: (1) content validity is sufficient as sole evidence supporting test scores interpretation, (2) evidence can be based on the relationship between content and construct but support from other types of evidence is needed, and (3) evaluating representativeness of construct by test content is an important process, but it cannot satisfactorily provide validity evidence. As for the quantification techniques, there were four techniques proposed by different authors. Each focused on a different aspect of the relationship between construct and test content, such as whether or not test content is relevant (e.g., Aiken's V and Polit's CVI), essential (Lawshe's CVR), or discriminant (Dixon and Johnston's Discriminant Content Validity) to another construct. All but one of the techniques were variations of measures

of agreement between experts regarding how well an item represented a construct. Only one technique used statistics comparing means of judgments on the content relevance to different constructs. Most techniques have been evaluated and revised or corrected by other authors, while Dixon and Johnston's DCV has not because it was just published currently. The impact of the findings was discussed.

Introduction

Validity is a crucial psychometric quality of a test related to the accuracy of test score interpretation (American Educational Research Association et al., 2014). The concept of validity has evolved from concepts emphasizing correlations of test scores with other variables, tripartite conceptualization, to the unitarian conceptualization of validity. In the unitarian conceptualization, the three types of validity in tripartite conceptualization did not stand by themselves as a discrete concept of validity but rather as classes of evidence to be collected in order to establish the validity of test score interpretation (American Educational Research Association et al., 2014; Santoso, 2010; Supratiknya, 2014).

One type of evidence that invites a lot of controversy was validity evidence based on test content or known as content validity in tripartite conceptualization. The controversion of content-based validity evidence happened not only at the conceptual level but also at the practical level, particularly on quantification techniques.

At the conceptual level, the position of content-based validity evidence as the validity evidence for test scores interpretation was often questioned (for example Beckstead, 2009; Guion, 1977; Messick, 1993). The authors rejected the content-based validity evidence as the evidence for validity because the definition of validity was related to test score interpretation while content-based validity provided evidence of the representativeness of test domain by test content. The evidence for validity can only be provided by analyzing test scores obtained from field testing, that is by comparing the behavior of test scores from the field testing with the behavior expected or hypothesized from the underlying theory.

Several other authors defended the position of content-based validity as one of the validity evidence (Chalhoub-Deville, 2009; Lissitz & Samuelsen, 2007a, 2007b). Content-based validity was considered important in the construction or development of tests with very specific domains such as competence or achievement tests. The validity of test scores interpretation could not be separated from the meaning and representativeness of test content. The class of evidence was also included in the recent *Standards* (American Educational Research Association et al., 2014). However, there was not any argument made in the *Standards* for maintaining content-based evidence as a class of validity evidence.

Unfortunately, discourses on the controversies were rarely reviewed and published. For example, books on psychological assessment and testing (Gregory, 2015; Urbina, 2014) did not include such discussion. As a result, knowledge of the evolution of validity conceptualization, particularly of content-based validity, did not reach a bigger audience in psychology. For that reason, the current study was conducted to review and then present the position of content-based validity hitherto.

At the practical level, many authors had proposed techniques for quantifying judgments from experts on content-based validity (Aiken, 1980; Johnston et al., 2014; Lawshe, 1975; Polit & Beck, 2006). The techniques were proposed because the effort of collecting evidence hitherto was considered superficial (Supratiknya, 2016) or subjective (Crocker & Algina, 2006).

However, to the current authors' knowledge, there was not any publication comparing and evaluating the techniques to find their strengths and limitations. Such knowledge could be used by test developers to choose appropriate techniques. For that reason, the current study also addressed issues related to the strength and limitations of quantification techniques.

Based on the presentation above, there were three questions addressed in the current study:
1. How was the position of content-based validity evidence in the concept of validity?
2. How were the comparison and evaluation of techniques for quantifying validity evidence based on test content?

Methods

Research Design
The current study was a literature review in the measurement area that included content-based validity as the topic of the resources not only methods to provide evidence of the validity of test score interpretation. The literature included topics on techniques for quantifying content-based validity.

Inclusion and Exclusion Criteria
The inclusion criteria used to select resources to be reviewed were:
1. Empirical studies or article reviews on content-based validity evidence and/or its quantification techniques
2. Content-based validity or its quantification was written as the topic of the resources
3. The year of publication was not limited because the author of the current study wanted to capture all arguments supporting or rejecting content-based validity as validity evidence.

The only exclusion criterion was that the content-validity evidence was treated not as the subject matter but only as the method to provide evidence of validity.

Data Collection
The articles reviewed in the current study were searched using Bielefeld Academic Search Engine (Base) that were then collected by buying them from the publishers. The authors also searched the articles from the bibliography included in reviewed articles. The keyword used in the search engine was *content validity* or *validity based on test content.* The results from the search were then selected following the aforementioned inclusion and exclusion criteria.

Analysis
The selected publications were then analyzed using the narrative thematic technique in the form of conclusions and discussion of findings or presentation in the literature (Cronin et al., 2008; Higgins & Deeks, 2008; Paré & Kitsiou, 2017). Following the steps

proposed by Cronin et al. (2008), the narrative thematic analysis was conducted using the PQRS method consisting of Preview, Question, Read and Summarize.

Results

Description of Resources Reviewed

The authors used the keyword "content validity" in the title column of Bielefeld Academic Search Engine to search for the articles to be reviewed. As many as 1841 journal articles were found, but only 24 of them met the inclusion criteria. We also add five more articles acquired from reading the materials and three chapters from three books that discussed content validity or validity in general.

Out of 32 resources, 15 of them were published in journals or books dedicated to discussing measurement topics, while 15 were published in substantive journals. It was interesting to find that out of 15 substantive resources, 7 of them were published in journals with nursing topics, while 8 were published in psychology and 2 in education. The results show that the topic of content validity seemed also to be an important issue in nursing research literature. The distribution of journals published articles on content validity topic can be seen in Table 1.

It is also worth noting that although the resources were published in different areas, the discussion on content-based validity evidence was similar. The similarity means that concerns regarding content-based validity evidence were the same across areas and may therefore across types of tests or scales.

Table 1. Distribution of journals that published articles on content validity

Title of Journals	Number of Articles
Measurement	
Measurement and Evaluation in Counseling and Development	2
International Journal of Selection and Assessment	2
Applied Psychological Measurement	2
Journal of Educational Measurement	1
Educational and Psychological Measurement	5
Book chapters on content validity	3
Psychology	
Industrial and Organizational Psychology	1
Personnel Psychology	3
British Journal of Health Psychology	2
Psychological Review	1
Psychological Reports	1
Educational	

Title of Journals	Number of Articles
Educational Researcher	2
Nursing	
International Journal of Nursing Studies	1
Nursing Research	1
Contemporary Nurse	1
Research in Nursing & Health	2
International Journal of Nursing Terminologies and Classifications	1
Western Journal of Nursing Research	1
Total	**32**

Controversies of Content Validity

There were three positions on the issue of whether content-based validity evidence could be considered as validity evidence. The first one argued that content-based validity evidence, or content validity as its proponent called it, was a different and independent type of validity. It provided information regarding validity of test score interpretation and was sufficient by itself. For example, Lissitz and Samuelsen (2007a) argued that content validity was essential to the evaluation of a test that was constructed based on a clear definition of the content domain such as the one in educational tests. The content validity, together with reliability, was then evaluated based on internal information from the test. Other authors also had the same positions, for example, Schmidt (2012) and Chalhoub-Deville (2009).

Furthermore, Lissitz and Samuelsen (2007a) stated that the content validity evaluated the adequacy of the process of constructing the test. Such a process could be applied to the creation of tests measuring a clear domain of interest as well as a latent construct. It is important to note that Lissitz and Samuelsen rejected the unitary concept of validity proposed by Messick (1989) and established their own concept.

Another proponent, Ebel (1956), stated more strongly that content validity was the basis for establishing other types of validity. For example, concurrent-criterion validity could be established by evaluating the correlation between the test under investigation and another test measuring the same construct that has already been proven to be valid. The availability of another valid test could only be justified using content validity because the search for a valid test without content validity would make the effort circular. Such a need for content validity was also mentioned by Lennon (1956) in a more moderate manner.

However, it is worth noting that both Ebel and Lissitz & Samuelsen (2007a) ignored the fact that the evaluation of the internal structure of a test and participants' response to items were two kinds of evidence that could also be provided without the need of other tests, thus lessen the need for content validity. Both kinds of evidence could also be

obtained by evaluating internal information of the test under scrutiny (Embretson, 2007).

The second position can be seen as the direct opposite of the first position in that the authors rejected the representativeness of test content as validity evidence. The rejection did not mean to diminish the importance of the process to evaluate the connection between test content and the construct intended to be measured. It was only that such a process was not considered enough to provide support for test score interpretation.

Messick (1989, 1993) was the prominent author who promoted the idea of the unitarian validity concept, in which all other types of validity were treated as types of evidence supporting the validity of test scores interpretation. Messick's opinion on content validity was actually not new. Several other authors such as Mosier (1947), Cureton (1951), Loevinger (1965), Guion (1977, 1978), and Tenopyr (1977) have already asserted that the process of evaluating the representativeness of construct by the test content could not be regarded as evidence of validity. Such a process only provided a hypothesis that needed empirical testing to support it. A more recent publication on job performance measures by Murphy (2009) showed that there were weak relationships between results from the content-based validation process and other validation processes. Other more current publications that also had the same position were Beckstead (2009a, 2009b).

The other problem of treating content representativeness as evidence for validity was that the process participants intended to conduct to arrive at a certain response may not be the same as the actual process the participants carried out (Messick, 1989). For example, the influence of social desirability and other response styles on the way test takers responded to the stimulus might not be identified in the evaluation of test content.

Although the authors in the second position rejected content representation as evidence of validity, they still considered the process of test content evaluation against the test domain was still extremely important. The importance of the process was due to the influence and limits of the nature and the content coverage imposed on the interpretation of test scores supported by other evidence.

The third group of authors posited a more moderate manner on content-based validity evidence. The authors' position was a middle ground between the first and second positions. While acknowledging the role of content representativeness in providing valid evidence, they pointed out the need of other types of evidence to establish the validity of test score interpretation. For example, Lennon (1956) stated that content-based evidence was needed when there was not any "dependable criterion variable" that could be readily accessed and when correlations with concurrent or future criteria were not meaningful as indicators of validity (p. 297). However, Lennon also stated that evaluation of the representativeness of the content of the tests was not sufficient to establish validity because other information, that is the process the test takers employed to arrive at their response, was also needed.

Embretson (2007) also provided a moderate position regarding the content-based validity evidence. In the article, Embretson first showed the limitation of content validity in establishing the validity of test scores interpretation. For example, content-based

validity evidence did not provide any evidence that test takers employed the skills, knowledge, or processes supposedly used in responding to test items. However, Embretson then included content-based validity evidence in eleven categories of evidence, particularly in Test Specification, Item Design Principles, and Domain Structure. Embretson argued that such evidence was crucial because the emphasis of providing evidence of validity should be on the internal evidence not on nomological networks of constructs. Similar opinions were also held by other authors such as Sackett (2012) and also the new *Standards* (American Educational Research Association et al., 2014).

We concluded that the first position was indefensible for several reasons. First, the process of content validation could only provide hypotheses about the relationship between test content and the construct underlying the test. Next, the content validation process could not provide sufficient evidence for the process presumed to be employed by test takers to arrive at certain responses. Furthermore, construct under-representativeness and construct irrelevant variance could only be identified by analyzing the test or item scores, not by evaluating the representativeness of the test content.

The third position was appealing because it provided a middle ground that accommodated both the first and second positions. However, the authors of the current article could not find any strong and fundamental argument for incorporating content representativeness as one type of evidence supporting the validity of the test scores interpretation. The only argument that we could find was in Lennon (1956). However, the articulated reason for conducting content validity was considered obsolete because there have been many methods developed to provide evidence for validity without the need for external criteria since. Therefore, the current authors chose to side with the second position in which the process of evaluating content representativeness was considered crucial but could not be considered as valid evidence.

However, we also found that the arguments posed by each position regarding content-based validity evidence could not be separated from the conceptualization of validity in general. For example, all authors in positions two and three believed that validity was attached not to the test but to the test score interpretation, therefore it was the score of the test that should be treated as the source of information regarding validity. While the prominent authors in position one believed otherwise. Therefore, the evaluation of the content validity position should be re-reviewed in a broader scope covering the conceptualization of validity in general. In turn, the conceptualization of validity itself could not be separated from how the measurement was conceptualized, therefore the future study should also include the philosophy underlying the definition of measurement.

Quantification of Content Validity

Although the current authors have chosen to side with the second position, issues regarding the quantification of content representativeness were still relevant. The quantification methods could be used in the process of establishing the relationship between test content and the test's underlying construct.

As far as the current authors could find, there were four methods for quantifying the content-representativeness that were presented in Table 2. The methods evaluated different aspects of content-representativeness. Lawshe's CVR, for example, evaluated whether an item was 'essential' to represent the construct measured, while Polit et al.'s CVI (2007) emphasized the item's relevance to the test. Aiken's V (1980) did not clearly identify what aspect of representativeness it measured. The author only mentioned that the judge needed to evaluate the content validity of the test. The newest method proposed was the Discriminant Content Validity which evaluated how well an item represented the underlying construct and the extent to which test items represented a different construct.

It was important to mention here that all but one, the discriminant content validity, has already been evaluated by different authors and shown that there were some problems regarding their inference. Fortunately, those articles also provided the revision of the formula for either calculating the standard error of the statistics or drawing conclusions regarding the results. And based on the accuracy of the results from the revisions, the current authors strongly recommend their use.

Table 2. Comparison of quantification techniques for content-based validity

	Lawshe	**Aiken**	**Polit, Beck & Owen**	**Dixon & Johnston**
Source	Lawshe (1975)	Aiken (1980)	Polit, Beck & Owen (2007)	Johnston et al., (2014)
What is evaluated	"Essential"	"Content Validity"	"Relevance"	"Discrimination"
How is the quantification conducted	Proportion of the essential judgment	Proportion of the ratings mean	inter-rater agreement on relevance	differences of relevance across constructs
Inference*	unclear	Multinomial or approximation of Normal Distribution	Inference of Fleiss Kappa	Mean difference tests such as t-test or Wilcoxon
Revision	Wilson, et al. (2012); Baghestani, et al. (2019)	Penfield et al., (2009)	Beckstead (2009a)	NA

Note: *Methods of inference found in the original article.

Conclusion

The first conclusion drawn from the discussion above was that the current author tended to agree with the second position regarding the content-based validity evidence, that content representativeness should not be considered as validity evidence. The current authors acknowledged the importance of the evaluation content representativeness

based on the underlying construct therefore such a process should be still conducted in the explication processes or prior to collecting field testing data.

However, the current author realized that the discussion regarding content-based validity evidence could not be addressed comprehensively without addressing the validity conceptualization that in turn could not be separated from the conceptualization of measurement. Therefore, the current author encouraged such discussion in future studies to provide a more comprehensive perspective on content-based validity evidence.

Regarding the quantification methods, we concluded that the use of the methods could still be continued as the process of establishing relationships between test content and the underlying construct should still be practiced. However, one should use the more recent formula or the revision of the original formula to provide a more accurate estimate and inference.

References

Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, 40 (4), 955–959. https://doi.org/10.1177/001316448004000419

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*.

Baghestani, A. R., Ahmadi, F., Tanha, A., & Meshkat, M. (2019). Bayesian critical values for Lawshe's content validity ratio. *Measurement and Evaluation in Counselling and Development*, 52(1), 69-73. https://doi.org/10.1080/07481756.2017.1308227

Beckstead, J. W. (2009a). Content validity is naught. *International Journal of Nursing Studies*, *46*(9), 1274–1283. https://doi.org/10.1016/j.ijnurstu.2009.04.014

Beckstead, J. W. (2009b). On content validity: A response to Squires (2009). *International Journal of Nursing Studies*, *46*(9), 1286–1287. https://doi.org/10.1016/j.ijnurstu.2009.06.009

Benson, J. (1981). A redefinition of content validity. *Educational and Psychological Measurement, 41*(3), 793-802. https://doi.org/10.1177/001316448104100320

Chalhoub-Deville, M. (2009). Content validity considerations in language testing contexts. In R. W. Lissitz (Ed.), *The concept of validity: revisions, new directions, and applications* (pp. 241–263). Information Age Publishing.

Crocker, L., Llarbre, M., & Miller, M. D. (1988). The generalizability of content validity ratings. *Journal of Educational Measurement*, *25*(4), 287-299. https://doi.org/10.1111/j.1745-3984.1988.tb00309.x

Crocker, L., & Algina, J. (2006). Introduction to Classical and Modern Test Theory. Cengage Learning.

Cronin, P., Ryan, F., & Coughlan, M. (2008). Undertaking a literature review: a step-by-step approach. *British Journal of Nursing*, *17*(1), 38–43. https://doi.org/10.12968/bjon.2008.17.1.28059

Cureton, E. E. (1951). Validity. In E. F. Lindquist (ed*.), Educational Measurement*. (pp. 621–694). American Council on Education.

Dixon, D., & Johnston, M. (2019). Content validity of measures of theoretical constructs in health psychology: Discriminant content validity is needed. *British Journal of Health Psychology, 24*, 477-484. https://doi.org/10.1111/bjhp.12373

Ebel, R. L. (1956). Obtaining and reporting evidence on content validity. *Educational and Psychological Measurement*, *16*(3), 269-282. https://doi.org/10.1177/001316445601600301

Embretson, S. E. (2007). Construct validity: A universal validity system or just another test evaluation procedure? *Educational Researcher, 36*(8), 449-455. https://doi.org/10.3102/0013189X07311600

Fitzpatrick, A. R. (1983). The meaning of content validity. *Applied Psychological Measurement*, 7(1),3-13. https://doi.org/10.1177/014662168300700102

Gregory, R. J. (2015). Psychological testing: History, principles, and applications (7th ed.). Pearson Education.

Guion, R. M. (1977). Content validity: The source of my discontent. *Applied Psychological Measurement, 1*(1), 1–10. https://doi.org/10.1177/014662167700100103

Guion, R. M. (1978). Content validity in moderation. *Personnel Psychology*, 31(2), 205-213. https://doi.org/10.1111/j.1744-6570.1978.tb00440.x

Higgins, J. P., & Deeks, J. J. (2008). Selecting studies and collecting data. In J. P. Higgins & S. Green (Eds.), Cochrane handbook for systematic reviews of interventions (pp. 151–85). John Wiley & Sons.

Johnston, M., Dixon, D., Hart, J., Glidewell, L., Schröder, C., & Pollard, B. (2014). Discriminant content validity: A quantitative methodology for assessing the content of theory-based measures, with illustrative applications. *British Journal of Health Psychology, 19*(2), 240–257. https://doi.org/10.1111/bjhp.12095

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, *28*(4), 563–575. https://doi.org/ 10.1111/j.1744-6570.1975.tb01393.x

Lennon, R. T. (1956). Assumptions underlying the use of content validity. *Educational and Psychological Measurement*, *16*(3), 294-304. https://doi.org/10.1177/001316445601600303

Lissitz, R. W., & Samuelsen, K. (2007a). Further clarification regarding validity and education. *Educational Researcher*, *36*(8), 482–484. https://doi.org/10.3102/0013189X07311612

Lissitz, R. W., & Samuelsen, K. (2007b). A suggested change in terminology and emphasis regarding validity and education. *Educational Researcher*, *36*(8), 437–448. https://doi.org/10.3102/0013189X07311286

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, *3*, 635-694. https://doi.org/10.2466/pr0.1957.3.3.635

Loevinger, J. (1965). Person and population as psychometric concepts. *Psychological Review, 72*, 143-155. https://doi.org/10.1037/h0021704

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, *35*(6), 382–386. https://doi.org/10.1097/00006199-198611000-00017

Messick, S. (1989). Validity. In L. R L (Ed.), *Educational Measurement* (3rd ed., pp. 13–103). American Council on Education; Macmillan.

Messick, S. (1993). Foundations of Validity: Meaning and Consequences in Psychological Assessment. *ETS Research Report Series*, *1993*(2), i–18. https://doi.org/10.1002/j.2333-8504.1993.tb01562.x

Michell, J. (2009). Invalidity in validity. In R. W. Lissitz (Ed.), *The concept of validity: Revisions, new directions, and applications* (pp. 111–133). Information Age Publishing.

Mosier, C. I. (1947). A critical examination of the concepts of face validity. *Educational and Psychological Measurement, 7*, 191–205. https://doi.org/10.1177/001316444700700201

Murphy, K. R. (2009). Content validation is useful for many things, but validity isn't one of them. *Industrial and Organizational Psychology*, *2*(4), 453-464. https://doi.org/10.1111/j.1754-9434.2009.01173.x

Paré, G., & Kitsiou, S. (2017). Methods for Literature. In F. Lau & C. Kuziemsky (Eds.), *Reviews Handbook of eHealth Evaluation: An Evidence-based Approach*. University of Victoria.

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, *29*(5), 489–497. https://doi.org/10.1002/nur.20147

Polit, D.F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health*, 30(4), 459-467. https://doi.org/10.1002/nur.20199

Sackett, P. R. (2012). Cognitive tests, constructs, and content validity: A commentary on Schmidt (2012). *International Journal of Selection and Assessment, 20*(1), 24-27. https://doi.org/ 10.1111/j.1468-2389.2012.00576.x

Santoso, A. (2010). Validitas: Dulu dan Sekarang. In A. Supratiknya & T. Susana (Eds.), *50 Tahun Himpunan Psikologi Indonesia: Redefinisi Psikologi Indonesia dalam Keberagaman*, (pp. 333–346). Himpunan Psikologi Indonesia (HIMPSI).

Schmidt, F. L. (2012). Cognitive tests used in selection can have content validity as well as criterion validity: A broader research review and implications for practice. *International Journal of Selection and Assessment*, *20*(1), 1-13. https://doi.org/10.1111/j.1468-2389.2012.00573.x

Supratiknya, A. (2014). *Pengukuran Psikologis*. Penerbit Universitas Sanata Dharma.

Supratiknya, A. (2016). *Kuantifikasi validitas isi dalam asesmen psikologis*. Penerbit Universitas Sanata Dharma.

Tenopyr, M. L (1977). Content-construct confusion. *Personnel Psychology, 30*(1), 47-54. https://doi.org/ 10.1111/j.1744-6570.1977.tb02320.x

Urbina, S. (2014). *Essentials of psychological testing*. John Wiley & Sons.

Wilson, F. R., Pan, W., & Schumsky, D. A. (2012, July). Recalculation of the critical values for Lawshe's content validity ratio. *Measurement and Evaluation in Counseling and Development*, *45*(3), 197–210. https://doi.org/10.1177/0748175612440286

Appendix

Web Application for Quantifying Content-Based Validity Evidence
https://mosaikstatlab.shinyapps.io/quant_conttent_validity/?_ga=2.36371104.1817675805.1658166107-1639260922.1658166107