Content Validity using Mixed Methods Approach:

Its application and development through the use of a Table of Specifications Methodology[1]

Isadore Newman
Florida International University

Janine Lim
Andrews University

Fernanda Pineda
Florida International University

Paper presented at the 2011 Annual Meeting of the American Evaluation Association, Anaheim, California, Nov.2-5, 2011.

Abstract

There is paucity of detailed information in the literature on how to develop procedures for estimating content validity (logical validity) to increase the trustworthiness of assessment instruments. This article presents three unique examples, interpretation, and application of tables of specifications (ToS) for estimating content validity. To have an acceptable estimate of content validity the ToS must also have estimates of reliability. The procedures presented -Lawshe's (1975) Content Validity Ratio and Content Validity Index, and expert agreement estimates procedures- would enhance both. The development and the logic of the ToS requires presenting evidence that has transparency and creates trustworthiness in the validity estimates by maintaining an audit trail and through the use of triangulation, expert debriefing, and peer review. An argument is presented that content validity requires a mixed methods approach since data are developed through both qualitative and quantitative methods that inform each other. This process is iterative and provides feedback on the effectiveness of the ToS through a process of determining consensus.

Introduction

While there has been much written on the topic of content validity (logical validity), there is paucity of information in the literature on how to develop procedures for estimating it.

[1] The authors would like to express their gratitude to Drs. John Hitchcock, Carolyn Ridenour and Carole Newman for their feedback on this paper.

Therefore, a need exists to explore methods for improving the estimation and trustworthiness of this measure of validity. Trustworthiness in this context is understood to be the transparency and the accumulation of evidence that supports the logical argument. Although there are many references to content validity it in the literature (Nunnally; 1978, Nunnally & Bernstein, 1994; Haynes, Richard, & Kubany, 1995; Ary, Jacobs, & Razavieh, 2005; Beggs & Lewis, 1975; Johnson & Christensen, 2004), there is relatively little comprehensive discussion on the details involved in developing content validity procedures (Johnson & Christensen, 2004, Kerlinger & Lee, 1999). Historically, content validity has been very important, but little has been done to improve its credibility and fidelity. In this era of accountability, it is essential to not only to provide transparency, but evidence that creates confidence in the assessment instrument. It seems logical that the more types of content validity estimates one has, the greater the trustworthiness of those estimates one can achieve. There is however no one methodological procedure that would adequately provide this type of evidence. The authors share Teddlie and Tashakkori's (2009) position that the most effective way of providing this multitude of evidence would be through a mixed methods approach, which requires the triangulation of several types of data.. They indicate that mixed methods may generically improve validity estimates (trustworthiness) (Teddlie & Tashakkori, 2009).

This article presents the use of a table of specifications (ToS) for estimating content validity, theoretically and in practice. The development and the logic of the ToS requires presenting evidence that has transparency to increase the trustworthiness of the validity estimates by maintaining an audit trail-triangulating multiple data sources, and by expert debriefing, and using peer review. The argument presented herein is that content validity requires the use of mixed methodologies since data are developed through both qualitative and quantitative methods

that inform each other. This process is iterative and provides feedback on the effectiveness of any ToS by calculating consensus of the responses that are generated by an appropriately selected panel of expert judges.

The techniques to estimate content (or logical) validity in this paper are presented through the three examples of tables of specifications. The estimation strategies presented could be useful for evaluators, teachers, administrators, supervisors, or for those with limited resources or those who have access to people with expertise. The paper is organized into four main sections. It opens with a brief review of the concept of content validity. The second section presents an argument that the process of using a ToS for estimating content validity is mixed methods research. This is followed by a description of and how to construct a table of specifications (third section) and three applications are provided. The authors conclude with remarks about the benefits of using a ToS in evaluation and research.

Content Validity

Content validity, which is also called "definition validity" and "logical validity" (Newman, Newman, Brown, & McNeely, 2006), estimates how representative instrument items are of the content or subject matter that instrument seeks to measure (p. 48). For each included item, the question should be asked "Does this item look like it measures the content I want it to measure?" Content validation is a multimethod process (Haynes, Richard, & Kubany, 1995). One could count, classify, rate (Guion, 1977) or simply "speak" consensus. Usually from a qualitative stance, content validity provides (oral) indication of consensus by experts in the content area at hand (Newman & McNeil, 1998). However, there are some strategies that can be used to actually quantify consensus (i.e., estimate content validity). An example is Lawshe's

(1975) content validity ratio (CVR) (Appendices B and C), that will be discussed later. Either from a qualitative or quantitative stance, content validity estimates might be subject to change with time, changes in the instrument's construct, changes in the audiences/populations (Haynes, Richard, & Kubany, 1995), and changes resulting from newer literature that may be pointed to as best practices. In other words, content validity has a dynamic nature (Haynes, Richard, & Kubany, 1995). A pragmatic approach to content validity -that is, by using mixed methodology- could prepare the researcher for such changes.

Content Validity and Mixed Methods Research Design

We argue that strategies for establishing content validity are well aligned if not based on mixed methods. Qualitative data are typically acquired through focus groups with experts or members of the community/audience using an instrument, and then transformed into numeric values, scores or rankings. This process becomes iterative as the experts provide feedback, the literature is reviewed, and consensus is sought (Newman, Newman, Brown, & McNeely, 2006; Haynes, Richard & Kubany, 1995; Cronbach, 1970; 1971).

This approach allows for gathering alternative assessment evidence that might not be as popular, well known, or may be considered too lax in an era of standardized accountability in which research is being held to higher standards. Validity and reliability estimates, as Messick (1994) suggests, "should be uniformly addressed for all assessments because they are not just measurements principles, they are *social values* that have meaning and force outside of measurement wherever evaluative judgments and decisions are made" (p. 13). Using mixed methods research, a panel of experts in a focus group (working session) could provide the data

that would allow ~~the~~ quantification of the consensus of those *social values* that are key to the audience at hand.

The Table of Specifications

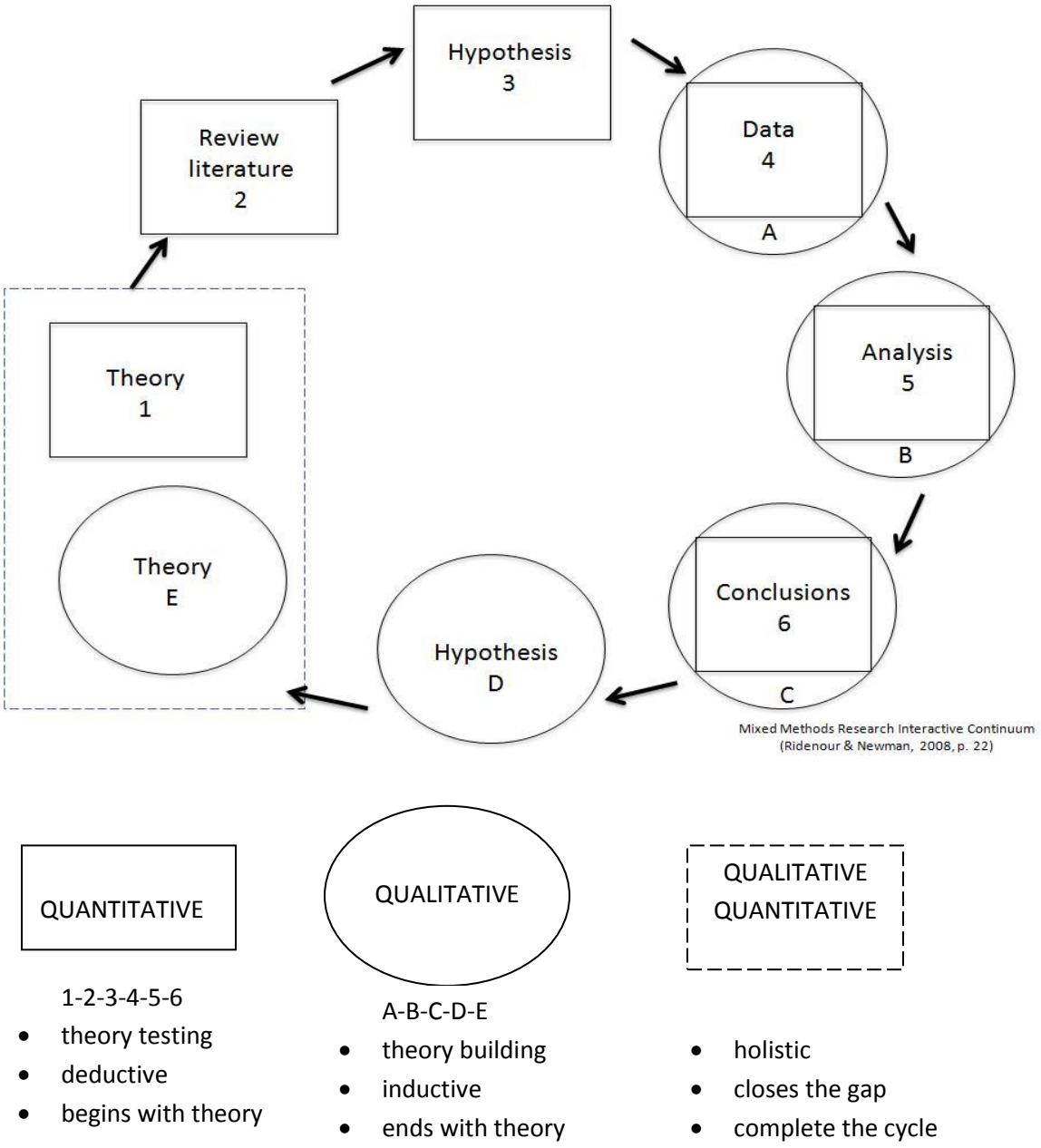
A Table of Specifications (ToS) is operationally defined by a set of procedures that attempt to align a set of items, tasks, or evidence with a set of concepts that are to be assessed. There are a variety of ways to gather evidence that estimates this alignment. These procedures can be quantitative, qualitative, or both. A ToS is a specific operational approach for the purpose of assessing the above intended alignment. An early operational definition provided by Newman, Frye, Blumefeld and Newman (1973), is the basis for our first example herein, the development of a test using Bloom's Taxonomy (Appendix A). Almost all references to a ToS tend to give a specific example. For example, in the work done by Fager (n/d), she demonstrates how to use a ToS to design a test, allocating the number of items to cover course materials. However, her use of the ToS appears to be more of an advanced organizer than as a tool for estimating content validity.

In reality, when talking about a ToS, one is generally referring to the alignment of any concept or construct to be assessed with the method to assess it. This paper demonstrates how to achieve that alignment and therefore it provides a more holistic approach to the concept and utilization of a ToS to estimate content validity. This is consistent with the following explanation of the underlying assumptions and operationalization of the process, indicating how mixed methods is the framework for this type of research.

The Table of Specifications and Mixed Methods Research Design

Our mixed methods approach to content validity is based on an inductive-deductive philosophy suggested by Ridenour and Newman (2008), which can be represented visually by two diagrams (figures 1 and 2).

Figure 1. The Qualitative-Quantitative Continuum of Research Methodology Conceptualized (adapted from Ridenour and Newman, 2008).



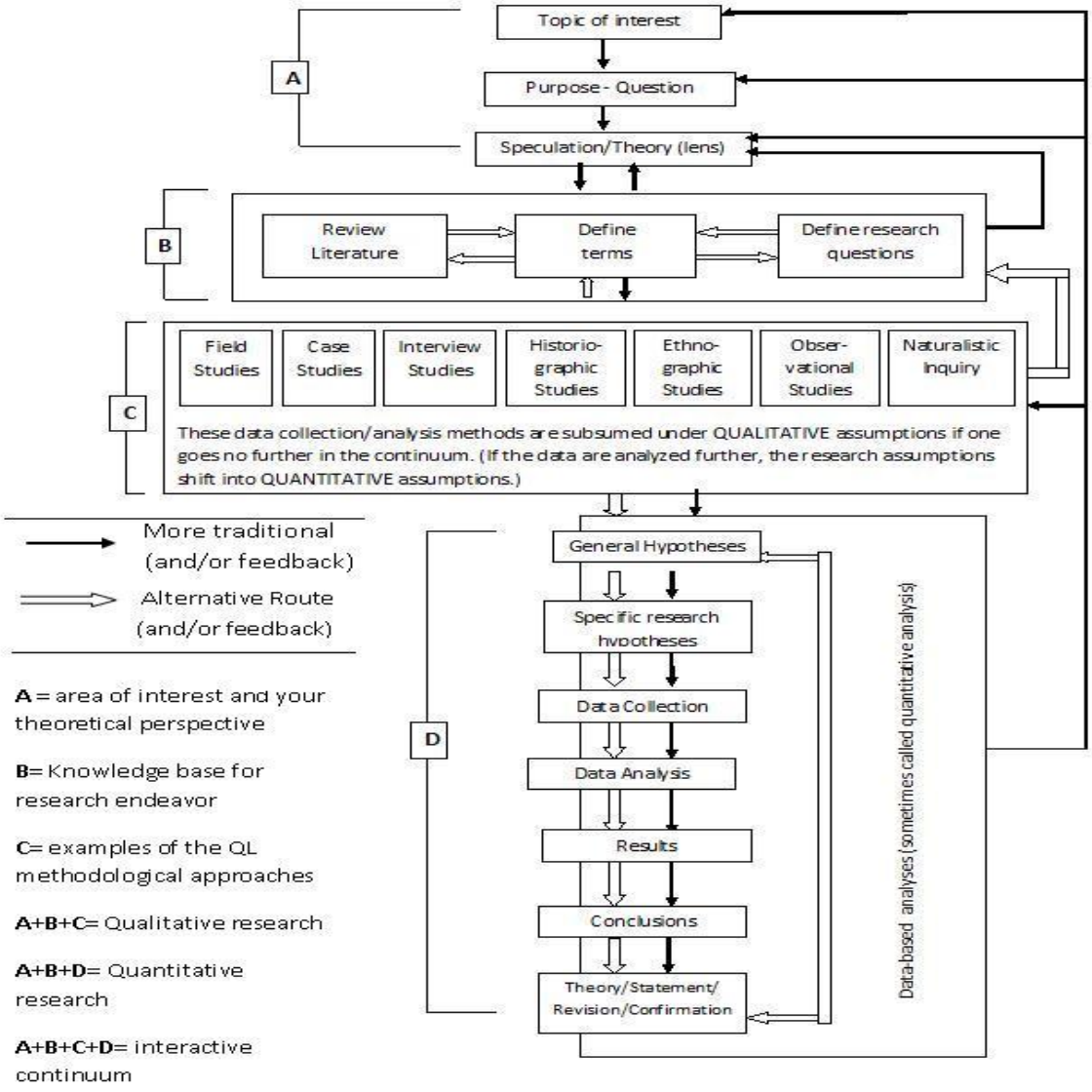
This diagram (figure 1) demonstrates the interactive continuum conceptualization in that, if one has a theory, by using a deductive approach you can then go from the theory to the review of the literature and the logic that supports the theory, which would lead to the production of hypotheses. You can then collect data to test the hypotheses, analyze the data, draw conclusions and then indicate if the data confirm the theory (numbers 1-6 in the diagram). The data being collected (#4, A), could be quantitative or qualitative. If the data is qualitative (A), it can be used heuristically to potentially give us insight into alternatives or building a more accurate theory.

Applying this model (figure 1) to the ToS procedures, the development of the concepts/constructs is conceptually the theory (#1) based upon the review of the literature (#2) and the experts. If these constructs are correct and well defined, then agreement (consensus) can be quantified by calculating the percentage of like responses from the appropriately selected experts, and by asking the judges: "Are these items on the ToS sufficient to measure this content area?" Also, judges can be encouraged to add what they think makes the assessment more representative of the construct (these procedures are discussed in a separate section). This aspect is qualitative, because it facilitates understanding of contexts, participant meaning, and can yield feedback from multiple perspectives.

In constructing the ToS we are collecting data, both qualitative and quantitative. We then analyze the data to get an estimate of how adequate the items are for assessing the concepts/constructs. Quantitatively, procedures include percentage of agreement (usually on the rows of the ToS, as shown in figure 3), percentage of agreement of the representativeness of content (usually on the columns of the ToS, as shown in figure 3), potentially cut scores, and the scores that differentiate between known groups (predictive/discriminate validity). Qualitative data is also collected at the same time when experts suggest what additional items should be

added, and (B) by identifying potential themes from the experts' feedback. This process yields both quantitative and qualitative conclusions about the potential representativeness of the items. Based upon the conclusions drawn from the qualitative analyses, one may develop additional hypotheses to gather the data that may be needed next (D); which could result in a modified theoretical base (E). The diagram presented in figure 1 is indicative of both an inductive and deductive approach for estimating content validity.

Figure 2. Interactive Continuum (adapted from Ridenour and Newman, 2008)



Similarly, figure 2 is conceptually doing the same thing where A, B and C may be considered qualitative research (inductive and theory producing) and A, B, and D may be considered quantitative research (more likely to be deductive and theory testing). When you are including A, B, C, and D, you then have an interactive continuum, where you have qualitative and quantitative methods informing each other, which is more representative of science (Ridenour & Newman, 2008). This model provides a frame for developing content validity estimates. For example, one starts off with "topics" (part A in figure 2) but the purpose/research questions inform each other to better delineate the topic of interest to make it more effective to be investigated (that is, moving effectively to section B in figure 2). The topic of interest and the purpose are informed by speculation and theory, which may be informed by the review of literature (section B in figure 2). Once A has informed B, and B has informed A (that iteration has occurred), one can employ methods (section C) that can be used to assess questions that emerge from part A,and that would be Qualitative research (A, B, and C). In quantitative research one uses A, and based upon A, develops research hypotheses that are testable. Data are then collected, analyzed, and conclusions are drawn based on the analysis of the data. The researcher then infers back to the data that either supports, disconfirms, or indicates a need to modify the theory. Steps A, B and D encompass the quantitative approach. In addition, this model has a feedback loop that allows the qualitative and quantitative data to inform the each other, which aligns with a mixed methods philosophy.

"A" relates to how the constructs/concepts in the ToS are identified. The hypothesis and data collection in the ToS are estimated by the rows and columns percentages (which would be part of "D"). The column data of the ToS also has a qualitative part, which is based upon subjective (phenomenological) experiences that suggest how well these items represent the

categories and also suggests other items that may facilitate the estimates of the constructs (part C in figure 2).

How to Construct and Use a Table of Specifications

The ToS has a structure such that the columns generally list the concepts, the topics of interest, the content areas to be covered by a certain program, and/or levels of achievement. The rows generally consist of the sources of evidence (items, artifacts, objectives); the items attempted to measure each concept are specified in those rows (Newman & McNeil, 1998; McNeil, Newman, & Steinhauser, 2005). The row items attempt to estimate the alignment between the rows and the columns in a number of ways:

A) Each judge can be asked to indicate if a row source is an estimate of a specific column topic or concept. For example, if the row evidence is "jumping rope" and the concept (column) is "Physical Coordination", the question that can be asked is, "Is jumping rope an estimate of having balance?" If you have 5 judges, and 4 agree, you would get 80% agreement that the item "jumping rope" is an estimate of the column concept of "Physical Coordination". If there were more row items such as "eye-hand coordination needed for a video game," "bouncing a ball," "juggling," and "hitting a baseball with a bat," then one can ask whether these items are sufficient for estimating "Physical Coordination".

B) Each judge can then be asked to rate the sufficiency of the checked items for estimating the column concept on a scale from 0% to 100% (it could be in a yes/no format as well). If it is not rated as 100% sufficient, space should be provided for judges to write their feedback (open-ended) indicating what additional evidence should be added to make it more sufficient. The

ratings for each person can then be averaged across all judges and recorded on a master form (figure 4).

C) The feedback provided by the judges (open-ended question) would then be qualitatively analyzed to see if there are any relevant themes that emerged from the feedback that should be added to improve the estimate of the concept. This process is repeated for each of the concepts.

Steps A, B, and C are repeated for each item on the ToS. McNeil, Newman, and Steinhauser (2005) point out that it is unusual to find a set of items that will actually measure 100% of the content to be evaluated, so an evaluator/researcher should aim to measure at least 80%. In the validity and reliability literature, achieving 80% agreement is the rule of thumb for having confidence in the instrument when estimating face, content and expert judge validity (Newman, Newman, & Newman, 2011).

There should be two ToS forms developed: one to give to each judge (figure 3), and the second to be used for master compiling of the expert judges' responses (figure 4) by the researcher/evaluator.

Figure 3. Expert Judge's Table of Specifications

Judge's ToS

Evidence	Concept 1		Concept 2		Concept 3
Item (artifact, objective) 1					
Item (artifact, objective) 2					
Item (artifact, objective) 3					

Item (artifact, objective) 4						
Item (artifact, objective) 5						
% to which the item estimates the concept:						
	Feedback* for Concept 1 (column)		Feedback for Concept 2 (column)		Feedback for Concept 3 (column)	

* The researcher can opt to have a space for feedback not only on "concepts", but on "items".

Figure 4. ToS Compiling Form used by the Researcher

Compiling Form - ToS

Evidence	Concept 1		Concept 2		Concept 3	% Agreement of Average of all Judges
Item (artifact, objective) 1						
Item (artifact, objective) 2						
Item (artifact, objective) 3						
Item (artifact, objective) 4						
Item (artifact, objective) 5						

Tally of Checkmarks; Sufficient? Yes/No							
% to which the item estimates the concept:							
	Themes for Concept 1 (column)		Themes for Concept 2 (column)		Themes for Concept 3 (column)		

- By looking at the marginal values (rows and columns) one can get an estimate of how well each concept is being estimated as well as which items of sources of evidence are sufficient or insufficient.

- In your master form, tally how many agreement checks there were for each concept, and then record it. This will give you the percentage of agreement between judges that the items measure the concepts of interest per concept.

ToS Applications

There can be many types of tables of specifications. One type of ToS can relate to selecting content to determine, "How much of each content area should be included?" Once you have identified the content, a ToS can help explore the items' representativeness of that content through the use of questions such as: "How well do the items represent this content area?" A ToS can also be constructed around the need to link achievement goals with evidence of that achievement (behaviors, descriptions of tasks, etc) with questions such as, "How well does the evidence represent the achievement of desired content, goals, or objectives?" A ToS can be constructed around all these concepts and many others. The following are examples of how to

develop a ToS that will support the validity of an assessment instrument or procedure. In this regard, the ToS is the major source of evidence for the content validity of the assessment device that one develops. We are suggesting two approaches for evaluating the data- one is a simple percentage calculation ("80% of the domain is perceived to be represented by the items on average") and the other one is Lawshe's Content Validity Ratio and Content Validity Index (Appendix B and C). The first approach is simpler and can be used effectively even with a small number of experts. Accepting 80% agreement is the rule of thumb regarding estimates of face, content, and expert judge validity (Newman, Newman, & Newman, 2011). The second approach should be considered with a larger number of experts.

The development of the ToS and the logic that goes into building it requires presenting evidence that is supportive of the transparency of the research procedure and the trustworthiness of the validity estimates. The development of a credible ToS provides evidence for validity while maintaining an audit trail for transparency, as well as for triangulation of data, expert debriefing, and peer review.

Application 1: **An Alignment of Content through the Support of Literature, Theory or Logic**

In a description of how to uses a ToS, Newman, Frye, Blumenfeld, and Newman (1973), provided the following examples of a ToS that was created to determine appropriate content as it relates (aligns) to theory, literature, and/or logic. The second example presented assesses item-content alignment on the ToS, which could be derived from the first ToS.

To estimate content validity for a test, a ToS was developed to determine if Bloom's Taxonomy levels are adequately represented, on the instrument. Newman et al. suggested that expert judges could be asked to answer the following questions, for the items (the columns) and

the descriptions of the performance or behaviors needed (the rows). For the Bloom's Taxonomy levels (this could also be content area, for example) one question could be, "Are the Bloom's Taxonomy descriptions in appropriate proportion?" Based on the expert judges' responses and the review of the literature, a conclusion could be drawn about the representative of the items, To determine which items on the test are appropriate, one could ask: "Are these items sufficient to measure the levels of Bloom's Taxonomy?" This ToS is available in Appendix A.

Application 2: Table of Specifications for the development of an instrument for K12 coordinators implementing curriculum videoconferencing

In this example, Lim (2010) first conducted a literature review to obtain the list of utilization predictors (i.e., the factors that, when present, yield higher utilization of videoconferencing). She then developed an instrument (a survey), using a ToS which aligned the survey items and the utilization predictors for videoconferencing in K12 education. An excerpt of this ToS is available in Appendix A.

Application 3: Table of Specifications for the creation of a Performance-Based Assessment Instrument (A Rubric)

The third example presented is Pineda's (2011), which serves as the basis for creating a performance-based assessment instrument (a holistic rubric) to evaluate the level of attainment of competencies at a higher education institution. (For the full description of this study, see Pineda, 2011.) This type of ToS is appropriate to use when there are identified outcomes and one is interested in what evidence can be assessed to estimate the attainment of those outcomes. In this case, the content is the outcomes (or competencies) and the items are really types of evidence. Since the institution has already established the outcomes (the five competencies), experts' consensus is not needed for those. However, there is a need for agreement on the evidence that is

appropriate for estimating the achievement of the outcomes. These achievements can be on a continuum as in a rubric ("beginning, developing, accomplished, and exemplary"). This ToS is available in Appendix A.

Usefulness of the ToS in Evaluation and Research

Since the late 1970s, participatory approaches to evaluations in the international development arena (Cullen, Coryn, & Rugh, 2011), as well as in national/local contexts, have been the cornerstone for practitioners. Including the active participation of as many stakeholders as possible through member-checking, is no longer an option but a key component to a transparent, trustworthy, and useful evaluation. This is undoubtedly challenging, but the use of a ToS facilitates this kind of participation. Member-checking is a trustworthiness measure that is conceptually very similar to the using a ToS which allows the evaluator to design products that are useful and congruent with stakeholders' interests, by inviting participation and valuing their imput.

Summary

This paper presented an argument about content validity having the underlying assumptions of mixed methodology, and operationalized this with the concept and use of the Table of Specifications. The estimating strategies presented here could be useful for evaluators, teachers, administrators, supervisors, or for those with limited resources or only access to people with expertise. This paper included a brief review of the concept of content validity, an argument that the process of using a ToS for estimating content validity is mixed methods research, followed by a description of and how to construct a table of specifications. Three applications were provided.

APPENDICES

Appendix A – Three Applications of Table of Specifications

Appendix B – Lawshe's CVR Calculations

Appendix C – Lawshe's CVR values (minimum) and their significance for one tailed tests

Appendix A

Example 1. Table of Specifications for developing a test, using Bloom's Taxonomy: Experts provide feedback on each row, and on each item for validity estimates' purposes (original image from Newman, Frye, Blumenfeld, and C. Newman, 1973). The number at the top of the cells indicates the objective, and the number at the bottom indicates the test item.

UNIT	CONTENT	KNOWLEDGE	COMPREHENSION	APPLICATION	ANALYSIS	SYN./EVAL.	AFFECTIVE	PSYCHOMOTOR	TOTALS
OXYGEN	History and occurrence	1,2,3 1, 2, 3, 4, 5, 6	4 7, 8, 9, 10, 11						4 11
	Preparation: Lab. and Commercial			5, 6 12, 13, 14, 15, 16	7 17, 18, 19	8 20, 21, 22			4 11
	Properties and Uses	9, 10 23, 24, 25, 26	27, 28	11, 12 29	13, 14 30, 31				
HYDROGEN	History and occurrence	15 32, 33, 34	16 35						2 4
	Activity series	17, 18 36, 37, 38, 39	19 40, 41				20 42, 43, 44, 45, 46		4 11
	Preparation: Lab. and Commercial	21 47	48	22 49, 50	23, 24 51, 52, 53			25	5 7
	Properties and Uses	26, 27, 28 54, 55, 56	29 57, 58, 59				30 60		5 7
		12 21	4 14	5 8	5 8	1 3	2 6	1	30 60

Example 2. Table of Specifications for the development of an instrument for K12 coordinators implementing curriculum videoconferencing (Lim, 2010)

This is an excerpt of the ToS developed to align survey items and the utilization predictors for implementing curriculum videoconferencing among K12 coordinators. For each row, experts place an X in the column(s) that they believe the item measures. Experts are given space for feedback towards the end. For the entire ToS for this study, please refer to Lim (2010) (full reference available below).

Survey Item **Directions:** For each row, please check the column(s) that you believe the item measures.	**Technology**	**Coordinator's Ability to Support VC**	**Curriculum Integration**	**Support for Coordinator**	**Working with Teachers**	**None of the above**
1. Where is the videoconference system located? a. mobile within one school b. mobile within more than one school c. fixed classroom d. media center/library e. computer lab f. conference room. g. other						
2. Location of the videoconference system. a. The current location/mobility of the VC unit works well in our building. b. The current location of our VC unit usually works but could be better. c. The current location/mobility of our VC unit makes it hard to use. d. We haven't decided yet which location is best for our VC unit. e. Not applicable						

Survey Item **Directions:** For each row, please check the column(s) that you believe the item measures.	**Technology**	**Coordinator's Ability to Support VC**	**Curriculum Integration**	**Support for Coordinator**	**Working with Teachers**	**None of the above**
3. What is the primary reason the videoconference system is located where it is? a. Technical reasons (wires, switches, networking, etc.) b. Proximity to coordinator c. Ease of use for teachers d. Only available room e. Other						

Example 3. Table of Specifications for the Development of a Performance-based Assessment Instrument (A Rubric) (Pineda, 2011)

This ToS, emerging from Pineda's (2011) research, allows for the creation of a rubric which will evaluate the level of attainment of five competencies at a higher education institution that recent graduates (in a focus group) deem pertinent. These students are recent graduates from the Bachelor's Degree of Intercultural Management for Development offered at an Intercultural University in Mexico. The five competencies have been accepted and documented in the program of study as the desired outcomes of the educational program.

The five competencies, worded as student learning outcomes, are:

Students will be able to:

1. Articulate scientific and indigenous knowledge and create initiatives that address both
2. Facilitate negotiation for and management of resources and information
3. Strengthen and design intercultural initiatives
4. Make visible regional (indigenous) knowledge and regional initiatives
5. Generate diagnostic and proposal-oriented knowledge

The rows below are the description of identifiable performance characteristics for each competency (evidence to support attainment of each competency), which change with the levels of attainment, i.e., "beginning", "developing", "accomplished", or "exemplary", which are the ToS' columns.

Experts indicate whether they agree with the description of attainment of each item by checking off the corresponding box, and to which percentage the descriptions are adequate. In this case, there will be 20 items (four levels of attainment per competency). The descriptions below (of Competency 1) are for illustration purposes only, since the recent graduates are the ones who will provide this information when the study takes place.

Description of Competency, Evidence and Level of Attainment	Beginning	Developing	Accomplished	Exemplary
Competency 1: "The student understands the concepts of scientific and indigenous knowledge in isolation, and does apply them into a community project"				
% to which the item estimates the level of achievement:				
FEEDBACK: (Please use these spaces to provide feedback)				
Competency 1: "The student differentiates between the concepts of scientific and indigenous knowledge, and is able to articulate initial ideas to use them both, even if in isolation of each other, in a community project"				
% to which the item estimates the level of achievement:				
FEEDBACK:				
Competency 1: "The student understands and differentiates between the concepts of scientific and indigenous knowledge, and is able to conceptualize and articulate ideas for their simultaneous use in a community project"				
% to which the item estimates the level of achievement:				
FEEDBACK:				

Competency 1: "The student understands, differentiates, and applies the concepts of scientific and indigenous knowledge, understands when they are in conflict and when they are complimentary, and is able to articulate and write proposals of community projects that apply both"				
% to which the item estimates the level of achievement:				
FEEDBACK:				
Tally of Checkmarks; Sufficient? Yes/No				

Appendix B

Lawshe's CVR Calculations

Another way of calculating content validity, or quantifying consensus, is using Lawshe's (1975) CVR. The formula is as follows:

$$\text{CVR}= \frac{n_e - N_{/2}}{N_{/2}}$$

n_e= the number of experts who agreed on the relevance of the item, behavior, or question

N= total members of the panel of expert judges

Total agreement will equal 1, the highest CVR value (Lawshe makes this to be .99 for ease of manipulation). A CVR is calculated for each item in the instrument.[2]

In the rubric example we had five experts, and four agree on the adequacy of an item (or an artifact or a behavior), the CVR is: $\frac{4-5/2}{5/2}$ = .6

[2] A point of interest in the literature is 80% agreement (4/5) is frequently used. However, as Lawshe indicated, the degree of agreement statistically is related to the number of judges (N). For example, 4 out of 5 will yield a Content Validity Ratio (CVR) of .64 which according to Lawshe is not high enough to meet the minimum CVR value to keep the item on the instrument even though there is 80% agreement; however, if you have 20 judges out of 30 agreeing, the minimum value to keep an item is .33, which is much less than 80% agreement. As one can see, Lawshe's emphasizes on the number of judges which is based upon the concept of a sampling distribution and the number of subjects in the sample to represent a population value. This is generally not considered in qualitative research or in the concept of saturation. It is not uncommon that QL researchers will say that after two or three additional samples generally not exceeding 12 to 16 subjects (you can cite Guest, Bunce and Johnson here) if no new themes emerge, the themes are assumed to be representative of an entire population. According to Lawshe's table (Appendix C), this is probably an underestimate of the number of subject one should look at before assuming saturation (or do you mean agreement? I am not sure I would equate the two).

The CVR estimate serves as reference point for another standard (metric) that one can use to estimate their subjective estimate of minimum goodness (I don't understand this). If the practitioner has access to a large number of subjects such as 20 or 30, the CVR offers a more precise statistical estimate of accuracy. For example, if you have 9 judges and 8 agree on an item, you get a CVR of .78, which is close to 80% cut score of agreement that one finds frequently in the literature. Hence, the researcher could choose to estimate validity using the CVR should she have a large number of participants.

Appendix C

Lawshe's CVR values (minimum) and their significance for one tailed tests

When all experts (or "panelists") say the description is appropriate, the computed CVR is 1. When it is more than half of them, but less than all, the CVR goes between zero and .99

Number of Panelists	Minimum Acceptable CVR Value
5	.99
6	.99
7	.99
8	.75
9	.78
10	.62
15	.49
20	.42
25	.37
30	.33

References

Ary, D., Jacobs, L. and A. Razavieh (2005). *Introduction to Research in Education* (7th ed). New York, NY: Wadsworth Publishing.

Beggs, D. and E. Lewis (1975). *Measurement and evaluation in the schools*. Boston, MA: Houghton Mifflin Company.

Cronbach, L. J. (1970). *Essentials of psychological testing* (3rd ed). New York, NY: Harper & Row.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (ed). *Educational measurement* (2nd ed). Washington, DC: American Council on Education.

Cullen, A., Coryn, C., and J. Rugh (2011). The politics and consequences of including stakeholders in international development evaluation. *American Journal of Evaluation, 32*(3), 345-361.

Fager, J. (n/d). *Designing a test: Developing a table of specifications*. Retrieved from: https://academicaffairs.cmich.edu/caa/assessment/posters/tableofspecifications.pdf

Guion, R. (1977). Content validity: The source of my discontent. *Applied Psychological Measurement, 1*(1), 1-10.

Haynes, S. N., Richard, D. C. S. and E. S. Kubany (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7*(3), 238-247.

Johnson, B. and L. Christensen (2004). *Educational research: Quantitative, qualitative, and mixed approaches* (2nd ed). New York, NY: Pearson.

Kerlinger, F. N., and H. B. Lee (1999) *Foundations of behavioral research* (4th Ed). Wadsworth Publishing.

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

Lim, J. (2010). Conducting research section C: Table of specifications. Retrieved from http://www.janinelim.com/leadportfolio/pages/4bConductingResearch.htm.

McNeil, K., Newman, I., and J. Steinhauser (2005). *How to be involved in program evaluation: what every administrator needs to know*. Lanham, MD: ScarecrowEducation.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2), 13-23.

Newman, I., Frye, B., Blumenfeld, G. and C. Newman (1973). *An introduction to the basic concepts of measurement and evaluation*. Akron, OH: The University of Akron.

Newman, I. and K. McNeil (1998). *Conducting survey research in the social sciences*. New York, NY: University Press of America.

Newman, I., Newman, C., Brown, R. & S. McNeely (2006). *Conceptual statistics for beginners* (3rd ed). New York, NY: University Press of America.

Newman, I., Newman, D. and C. Newman (2011). Writing Research Articles Using Mixed Methods: Methodological Considerations to Help You Get Published. In T. S. Rocco and T. Hatcher (Eds). *The Handbook of Scholarly Writing and Publishing* (pp. 191-208). San Francisco, CA: Jossey-Bass.

Nunnally, J. (1978). *Psychometric theory* (2nd ed). New York, NY: McGraw-Hill Book Company.

Nunnally, J. and I. Bernstein (1994). *Psychometric theory* (3rd ed). New York, NY: McGraw-Hill Book Company.

Pineda, M. F. (2011). Unpublished dissertation proposal: "*The Development and Validation of an instrument to measure the attainment of competencies at an intercultural university in Mexico*". Miami, FL: Florida International University.

Ridenour, C. and I. Newman (2008). *Mixed methods research: Exploring the interactive continuum*. Carbondale, IL: Southern Illinois University Press.

Teddlie, C. and Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Los Angeles, CA: Sage Publications.