Paper 255-27

Computing Indices of Item Congruence for Test Development Validity Assessments

Ronna C. Turner, University of Arkansas, Fayetteville, AR
Sean W. Mulvenon, University of Arkansas, Fayetteville, AR
Shawn P. Thomas, Pharmaceutical Product Developments, Inc.
Richard S. Balkin, University of Arkansas, Fayetteville, AR

ABSTRACT

The following provides a macro for test developers to assess content validity using the index of item-objective congruence measure (Rovinelli & Hambleton, 1977). The target audience for this information focuses on researchers and practitioners involved with the development of measurement instruments. The fields of expertise could range from higher education to business and government. SAS/IML® is used to provide the classical unidimensional measure developed by Rovinelli and Hambleton (1977). This measure is limited to items that are measuring a single construct or a specific composite of constructs. In modern test theory, it is common to develop items that have multiple assessment targets. Thus, a macro of a newly developed index for evaluating items that measure multiple objectives or combinations of constructs is also provided in the paper.

INTRODUCTION

The development of measurement instruments is a common process in almost all professional fields including education, psychology, medicine, business, and government. The instruments include measures of perceptions, attitudes, opinions, personality traits, managerial styles, psychological characteristics, creativity levels, basic skills, and satisfaction. The quality of the information obtained from these assessments depends on a common set of test theory procedures that are used to aid in the development of an instrument. The validity of the assessment information obtained from the administration of an instrument is highly connected to the rigor and appropriateness of the procedures in the developmental stages.

The development of measurement instruments is a process which includes both a) the development of the item and subscale components and b) the qualitative and quantitative assessments of the item and subscale parameters. In order to appropriately use and interpret data obtained from a measurement instrument, there must be operational definitions of the constructs being measured and information on the reliability and validity of the scores. This information assists users in placing appropriate meaning to the results obtained and interpreting the scores within the confines of the assessment parameters identified. In validity assessment, the three general types most commonly discussed are content, criterion, and construct validity. During the developmental stage of creating items from a table of test specifications, it is important that content validity is assessed. The creation of items using a table of test specifications as a blueprint is not evidence of content validity (Crocker & Algina, 1986). Evidence of content validity can be obtained from an evaluation, conducted by an independent expert panel, of the effectiveness of items in measuring one or more objectives. Berk (1984) stated that an evaluation of the congruence between items and objectives is the most important assessment during the content validation stage. If there is insufficient evidence that the items are measuring what they are intended to measure, the remaining item analyses are useless. An efficient measure for numerically assessing content experts' evaluations of items is the index of item-objective congruence (Rovinelli & Hambleton, 1977), in which ratings from content specialists are obtained in order to evaluate the match between test items and the table of specifications (Berk, 1984).

INDEX OF ITEM-OBJECTIVE CONGRUENCE

An evaluation using the index of item-objective congruence (Rovinelli & Hambleton, 1977) is a process where content experts rate individual items on the degree to which they do or do not measure specific objectives listed by the test developer. More specifically, a content expert will evaluate each item by giving the item a rating of 1 (for clearly measuring), -1 (clearly not measuring), or 0 (degree to which it measures the content area is unclear) for each objectives. The experts are not told which constructs the individual items are intended to measure, thus they remain independent and unbiased evaluators. For example, consider the following items selected from a math test. The test developer would place the items in a list as seen in Figure 1 with the different constructs that the items might be measuring provided in the columns. Then a copy of the list would be provided to each expert and they would assess the degree to which each objective was being assessed by each item using the rubric previously specified. If item 1 is intended to only measure algebra and item 2 is designed to measure geometry, an ideal set of ratings from a content expert would be as follows:

	Algebra	Geometry	Calculus
Solve for x. 3x + 2 = 5	1	-1	-1
How many degrees are in a right angle?	-1	1	-1

Figure 1. Ratings of items on three math objectives.

After the experts complete an evaluation of the items, the ratings are combined to provide indices of item-objective congruence measures for each item on each objective. The range of the index score for an item is -1 to 1 where a value of 1 indicates that all experts agree that the item is clearly measuring only the objective that it is hypothesized to measure and is clearly not measuring any other objective. A value of -1 would indicate that the experts believe the item is measuring all objectives that it was not defined to measure and is not measuring the hypothesized objective. The formula developed by Rovinelli and Hambleton is used under the assumption that there is only one valid objective being measured by each item. If an item is measuring multiple objectives, then the index value would be less than one. The premise of the index is to have high positive values on the objective an item is intended to measure and values close to -1 on all of the remaining objectives. The index of item-objective congruence developed by Rovinelli and Hambleton (1977) is computed using the equation:

$$I_{ik} = \frac{(N-1)\sum_{j=1}^{n} X_{ijk} - \sum_{i=1}^{N}\sum_{j=1}^{n} X_{ijk} + \sum_{j=1}^{n} X_{ijk}}{2(N-1)n}$$

where I_{ik} is the index of item-objective congruence for item k on objective i,
N is the number of objectives (i = 1, 2, …, N),
n is the number of content specialists (j = 1, 2, …, n), and
X_{ijk} is the rating (1, 0, -1) of item k as a measure of objective i by content specialist j (Berk, 1984, p.209).

Many other mathematical assessments used in assessing test reliability and validity such as Cronbach's alpha, biserial and Pearson correlations, factor analysis, and item response parameter estimation have their mathematical models as standard protocols in statistical software available to educators and business professionals. The index of item-objective congruence model is not one that is readily available in SAS, and because of the iterative nature of the calculations, the creation of an index of item-objective congruence variable using standard SAS® equations (listed after an input statement) does not work. The index of item-objective congruence is easy to compute using matrix algebra, thus SAS/IML was used to create a macro for providing this assessment tool.

The following is a SAS macro for computing the unidimensional index of item-objective congruence. There are three places where the user will be required to enter information into the program: 1) entry of data into a traditional SAS data set, 2) identification of the valid and invalid objectives for each item, and 3) entry of the number of items to be assessed at the end of the macro.

```
/* THE USER INPUTS DATA INTO A REGULAR SAS
DATASET PROVIDING A VARIABLE FOR THE RATER, THE
ITEM, AND THE RATING FOR EACH ITEM ON EVERY
CONSTRUCT.

THE DATA BELOW HAS BEEN ENTERED USING THE FORMAT
SHEET SUGGESTED IN FIGURE 1. RESPONSES TO A SET
OF QUESTIONS CAN BE ENTERED ONE RATER AT A TIME.
THE FOLLOWING DATA HAS 4 RATERS ASSESSING 4
ITEMS MEASURING ONE OF 5 CONSTRUCTS. */

   options ls=75 ps=70;
   data one;
   input rater item c1 c2 c3 c4 c5;
   cards;
   1 1  1 -1 -1 -1 -1
   1 2  1  0 -1 -1 -1
   1 3 -1 -1  1 -1 -1
   1 4 -1 -1  1 -1 -1
   2 1  1 -1 -1 -1 -1
   2 2  1 -1 -1 -1 -1
   2 3 -1 -1  1 -1 -1
   2 4 -1 -1  1 -1 -1
   3 1  1 -1 -1 -1 -1
   3 2  1 -1 -1 -1 -1
   3 3 -1 -1  1 -1 -1
   3 4 -1 -1  1 -1 -1
   4 1  1 -1 -1 -1 -1
   4 2  1 -1 -1 -1 -1
   4 3  1 -1 -1 -1 -1
   4 4  1 -1  0 -1 -1
   ;
   proc sort; by item; run;
   proc iml;

/*           USER INPUT REQUIRED                   */

/**********************************************
IDENTIFY WHICH CONSTRUCTS ARE VALID FOR EACH
ITEM. V1 REPRESENTS CONSTRUCTS FOR ITEM 1, V2
FOR ITEM 2, ETC... 1 = VALID CONSTRUCT AND 0 =
INVALID CONSTRUCT. FOR EXAMPLE, ITEMS 1 AND 2
ARE IDENTIFIED AS MEASURES OF CONSTRUCTS 1 (AND
NOT MEASURES OF CONSTRUCTS 2,3,4, AND 5). ITEMS
3 AND 4 ARE MEASURES OF CONSTRUCT 3, ONLY.
**********************************************/

   V1 = {1 0 0 0 0};
   V2 = {1 0 0 0 0};
   V3 = {0 0 1 0 0};
   V4 = {0 0 1 0 0};

/*           USER INPUT REQUIRED                   */

/* IN THE USE AND READ STATEMENTS THE USER MUST
SPECIFY THE NUMBER OF CONSTRUCTS TO BE
EVALUATED: FOR EXAMPLE C1, C2, C3, C4, C5 FOR 5
OBJECTIVES */

/* SPLITTING DATA INTO ITEM LEVEL SUBSETS */

   %macro itemcong(numitem);
   %do item = 1 %to &numitem;
     USE one VAR{item c1 c2 c3 c4 c5};
     READ all VAR {c1 c2 c3 c4 c5}
       where (item=&item) into I&item;
     close one;
   %end;

/* COMPUTING INDEX OF ITEM CONGRUENCE FOR EACH
ITEM */

   %do item = 1 %to &numitem;
   N&item=ncol(I&item);
   p&item = V&item[,+];  r&item =
     nrow(I&item);
   A&item = (V&item*I&item`)[1,+];
   B&item = ((-1*(V&item-1))*I&item`)[1,+];
   Avg&item = I&item(|+,|)/(nrow(I&item));
   Index&item = (((N&item)*A&item)-
      (B&item+A&item))/(2*(N&item-1)*r&item);
   %end;
   %do item = 1 %to &numitem;
   print "Item:" &item[format=2.0] "Index of
      Item Congruence:" Index&item[format=6.2];
   print "Valid Constructs:"
      V&item[format=6.0];
   print "Construct Mean:"
      Avg&item[format=6.2];
   %end;
   %mend itemcong;

/*           USER INPUT REQUIRED                   */

/**********************************************
ADD IN THE NUMBER OF ITEMS BEING ASSESSED IN THE
PARENTHESES: THERE ARE 4 ITEMS IN THIS EXAMPLE
**********************************************/

   %itemcong(4);
   run;
```

The type of output received includes the item number, the index value, a string identifying the correct set of objectives being measured and the average ratings of the experts on each objective.

```
Item: 1   Index of Item Congruence:   1.00
Valid Constructs:  1     0     0     0     0
Construct Mean: 1.00 -1.00 -1.00 -1.00 -1.00

Item: 2   Index of Item Congruence:   0.97
Valid Constructs:  1     0     0     0     0
Construct Mean: 1.00 -0.75 -1.00 -1.00 -1.00

Item: 3   Index of Item Congruence:   0.69
Valid Constructs:  0     0     1     0     0
Construct Mean:-0.50 -1.00  0.50 -1.00 -1.00

Item: 4   Index of Item Congruence:   0.81
Valid Constructs:  0     0     1     0     0
Construct Mean:-0.50 -1.00  0.75 -1.00 -1.00
```

The index value of 1 for item 1 indicates that all experts agreed that the item is clearly measuring objective 1 and clearly not measuring objectives 2 through 5, as hypothesized by the test developer. Item 2 also has a high index of item-objective congruence value (I_{jk} = .97), indicating that the experts agree that item 2 is clearly measuring objective 1 and not measuring objectives 3, 4 and 5. Only one of four experts believed that it is unclear whether the item is a measure of objective 2. Item 3 has a lower index value than the other items and there are two problems with the item that need to be addressed. First, the experts do not clearly agree that the item is measuring objective 3 (the valid objective). Either one half of the experts are unsure, or one of the experts believes the item is clearly not measuring objective 3. Additionally, there are experts that believe item 3 may be measuring objective 1. Follow-up evaluations of this item would be needed. The item may need to be reworded. The content experts rated item 4 as being a fairly clear measure of objective 3, and clearly not a measure of objectives 2, 4, and 5. There is slight uncertainty to whether or not this item measures objective 1 (invalid objective). Typically items 1, 2, and 4 would fall within the acceptable range of values for an analysis using 4 content experts.

INDEX OF ITEM-OBJECTIVE CONGRUENCE FOR MULTIPLE OBJECTIVES

Additionally, the intent of this paper is to go one step further. The index of item-objective congruence has been defined to be an assessment tool for evaluating items that measure only a singular objective (or one composite of objectives). Commonly, though, items are either purposely or practically multidimensional in nature, or are identified as being a measure of multiple objectives or construct areas. Consider the following example in which a researcher would have a need for a multidimensional content validity assessment. A psychologist has experimented with four identified counseling procedures for working with clients who are severely depressed. He has evidence that the use of certain combinations of treatments is more effective than singular treatments in reducing depression levels of clients. Thus, he wants to create training tapes of psychologist-client interactions where only the identified combinations of treatment procedures are used. The index of item congruence procedure would be extremely difficult to use in this situation because of the many different types of combinations that may be possible. Turner and Carlson (2002) provided an adaptation of the Rovinelli and Hambleton index to allow for the evaluation of items measuring more than one valid objective. Thus, the second purpose of this paper is to provide a SAS/IML macro of the adapted index (Turner & Carlson, 2002) for researchers working with multidimensional data. The development of the model and guidelines for assessment criteria can be reviewed in more detail in Turner and Carlson (2002).

An example of a multiple objective test item is in Figure 2. To answer this item correctly, one would need knowledge of 1) slope-intercept form which would require training in coordinate geometry and 2) algebraic manipulation procedures. This item would be a valid measure of both geometry and algebra. A student would not need knowledge of calculus to obtain the correct answer, thus the item is rated -1 on the calculus objective. If the Rovinelli and Hambleton (1977) index is used in this situation, the value would be less than 1 and dependent on the number of total objectives and the number of valid objectives. The adjusted index developed by Turner and Carlson (2002) will provide an index value equal to 1 and is interpreted the same way as in the unidimensional measure.

	Algebra	Geometry	Calculus
Convert the equation to slope-intercept form: 3X + 2Y = 8	1	1	-1

Figure 2. Ratings for an item measuring multiple objectives.

The equation for the adjusted index is:

$$I_{ik} = \frac{(N + 2p - 2)\sum_{j=1}^{r} X_{ijk} - p_k \sum_{i=1}^{N}\sum_{j=1}^{r} X_{ijk}}{2(N-1)r * p_k}$$

where I_{ik} is the index of item-objective congruence for item k on a set of objectives i,
N is the total number of objectives (i = 1, 2, …, N),
r is the number of content specialists (j = 1, 2, …, r),
p_k is the number of valid objectives for item k, and
X_{ijk} is the rating (1, 0, -1) of item k as a measure of an objective i by content specialist j (Turner & Carlson, 2002).

In computing the index for multiple objectives, it is important to note that the subscript i represents a potential set of objectives rather than a single objective. In the mathematics example provided, both objectives 1 and 2 are valid (algebra and geometry, respectively) while objective 3 (calculus) is not valid for this item. Therefore, the first component of the equation to be calculated would be the product of N times the summation of the experts' ratings on the first two objectives. Further details on computations are provided in Turner and Carlson (2002).

The following is a SAS macro for computing the multidimensional index of item-objective congruence. This formula can be used for both the unidimensional and the multidimensional situations. The multidimensional index has the same range of values and is mathematically equivalent to the unidimensional measure for one valid objective.

```
/* THE USER INPUTS DATA INTO A SAS DATASET
PROVIDING A VARIABLE FOR THE RATER, THE ITEM,
AND THE RATING FOR EACH ITEM ON EVERY OBJECTIVE.
```

```
THE DATA BELOW HAS BEEN ENTERED USING THE FORMAT
SHEET SUGGESTED IN FIGURE 1. RESPONSES TO A SET
OF QUESTIONS CAN BE ENTERED ONE RATER AT A TIME.
THE FOLLOWING DATA HAS 4 RATERS ASSESSING 4
ITEMS ON EACH OF 5 CONSTRUCTS.

THE FIRST LINE OF CODE REPRESENTS RATER 1 ON
ITEM 1 WITH RATINGS OF 1, -1, 1, -1, AND -1 ON
THE 5 CONSTRUCTS  */

   options ls=75 ps=70;
   data one;
   input rater item c1 c2 c3 c4 c5;
   cards;
   1 1  1 -1  1 -1 -1
   1 2  1  0  1 -1 -1
   1 3  1 -1  1 -1 -1
   1 4  1 -1  1 -1 -1
   2 1  1 -1  1 -1 -1
   2 2  1 -1  1 -1 -1
   2 3  1 -1  1 -1 -1
   2 4  1 -1  1 -1 -1
   3 1  1 -1  1 -1 -1
   3 2  1 -1  1 -1 -1
   3 3  1 -1  1 -1 -1
   3 4  1 -1  1 -1 -1
   4 1  1 -1  1 -1 -1
   4 2  1 -1  1 -1 -1
   4 3  1 -1 -1 -1 -1
   4 4 -1  1  0 -1 -1
   ;

   proc sort; by item; run;
   proc iml;

/*           USER INPUT REQUIRED                */

/*********************************************
IDENTIFY WHICH CONSTRUCTS ARE VALID FOR EACH
ITEM. V1 REPRESENTS CONSTRUCTS FOR ITEM 1, V2
FOR ITEM 2, ETC... 1 = VALID CONSTRUCT AND 0 =
INVALID CONSTRUCT. FOR EXAMPLE, THE FOLLOWING
FOUR ITEMS ARE IDENTIFIED AS MEASURES OF
CONSTRUCTS 1 AND 3 (AND NOT MEASURES OF
CONSTRUCTS 2, 4, AND 5).
*********************************************/

   V1 = {1 0 1 0 0};
   V2 = {1 0 1 0 0};
   V3 = {1 0 1 0 0};
   V4 = {1 0 1 0 0};

/*           USER INPUT REQUIRED                */

/* IN THE USE AND READ STATEMENTS THE USER MUST
SPECIFY THE NUMBER OF CONSTRUCTS TO BE
EVALUATED: FOR EXAMPLE C1, C2, C3, C4, C5 FOR 5
OBJECTIVES  */

/* SPLITTING DATA INTO ITEM LEVEL SUBSETS */

   %macro itemcong(numitem);
   %do item = 1 %to &numitem;
     USE one VAR{item c1 c2 c3 c4 c5};
     READ all VAR {c1 c2 c3 c4 c5}
       where (item=&item) into I&item;
     close one;
   %end;

/* COMPUTING INDEX OF ITEM CONGRUENCE FOR EACH
ITEM */

   %do item = 1 %to &numitem;
   N&item=ncol(I&item);
   p&item = V&item[,+];  r&item =
     nrow(I&item);
   A&item = (V&item*I&item`)[1,+];
   B&item = ((-1*(V&item-1))*I&item`)[1,+];
   Avg&item = I&item(|+,|)/(nrow(I&item));
   Index&item = (((N&item+p&item-2)*A&item)-
      (p&item*B&item))/(2*(N&item-1)
      *r&item*p&item);
   %end;
   %do item = 1 %to &numitem;
   print "Item:" &item[format=2.0] "Index of
      Item Congruence:" Index&item[format=6.2];
   print "Valid Constructs:"
     V&item[format=6.0];
   print "Construct Mean:"
     Avg&item[format=6.2];
   %end;
   %mend itemcong;

/*           USER INPUT REQUIRED                */

/********************************************
ADD IN THE NUMBER OF ITEMS BEING ASSESSED IN THE
PARENTHESES: THERE ARE 4 ITEMS HERE
********************************************/

   %itemcong(4);
   run;
```

The following is the output from this macro:

```
Item: 1  Index of Item Congruence:   1.00
Valid Constructs:  1     0     1     0     0
Construct Mean: 1.00 -1.00  1.00 -1.00 -1.00

Item: 2  Index of Item Congruence:   0.97
Valid Constructs:  1     0     1     0     0
Construct Mean: 1.00 -0.75  1.00 -1.00 -1.00

Item: 3  Index of Item Congruence:   0.84
Valid Constructs:  1     0     1     0     0
Construct Mean: 1.00 -1.00  0.50 -1.00 -1.00

Item: 4  Index of Item Congruence:   0.70
Valid Constructs:  1     0     1     0     0
Construct Mean: 0.50 -0.50  0.75 -1.00 -1.00
```

The interpretation of the indices for the multiple objective measures are the same as that for the single objective index. The content experts rated item 1 as clearly measuring objectives 1 and 3, as hypothesized, and clearly not measuring objectives 2, 4, and 5. The content experts also agree that item 2 is a clear measure of the objectives 1 and 3. Item 3 is a clear measure of objective 1, and not a measure of objectives 2, 4, and 5. At least one content expert has questioned whether item 3 is a clear measure of objective 3. Item 4 is a relatively clear measure of objective 3, but not a clear measure of objective 1. Additionally, one or two of the experts indicated that item 4 might be measuring objective 2. A test developer would likely consider this information sufficient to proceed with items 1, 2, and 3 in the item analysis process, but may want to revise and reevaluate item 4 before proceeding.

CONCLUSION

The purpose of the macros developed in this paper are to make content validity assessment using the index of item-objective congruence quicker and easier to calculate for the average instrument developer. Again, the macro for the index of item-objective congruence for multiple objectives is mathematically equivalent to Rovinelli and Hambleton's index when the number of valid objectives being measured is one, thus can be used in both situations.

TRADEMARKS

REFERENCES

Berk, R. A. (1984). Conducting the item analysis. In R. A. Berk, Ed., *A guide to criterion-referenced test construction* (pp. 97-143). Baltimore, MD: The Johns Hopkins University Press.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* Orlando, FL: Harcourt Brace Jovanovich, Inc.

Hambleton, R. K. (1984). Validating the test scores. In R. A. Berk, Ed., *A guide to criterion-referenced test construction* (pp. 199-230). Baltimore, MD: The Johns Hopkins University Press.

Rovinelli, R. J., & Hambleton, R. K. On the use of content specialists in the assessment of criterion-referenced test item validity. *Dutch Journal of Educational Research, 1977, 2,* 49-60.

Turner, R. C., & Carlson, L. (2002). *Index of Item-Objective Congruence for Multiple Objective Measures.* Unpublished manuscript, University of Arkansas.

CONTACT INFORMATION

Contact the author at:

Ronna C. Turner, Ph.D.
University of Arkansas
302 West Avenue Annex
Fayetteville, AR 72701
(501) 575-5593
(501) 575-5185 (fax)
rcturner@uark.edu
orme.uark.edu