Modifying the C Index for Use With Holland Codes of Unequal Length

Donald E. Eggerth
Michael E. Andrew
Centers for Disease Control and Prevention
National Institute for Occupational Safety and Health

The concept of congruence between person and occupation lies at the core of Holland's (1997) theory of career types. The C index is arguably the best available method for comparing the congruence of two Holland code profiles. The C index reflects the theorized hexagonal structure of the Holland RIASEC model, is sensitive to code ordering, and is simple to calculate. However, the C index is formulated to only make comparisons between Holland code profiles three letters in length. Although this is consistent with the instrumentation and supporting materials developed by Holland and his colleagues, it is inconsistent with both the Strong Interest Inventory (SII) and the Occupational Information Network (O*NET), each of which assigns Holland codes of one to three letters. Consequently, the C index cannot be easily used with either the SII or the O*NET. Moreover, the authors argue that it is arbitrary to always calculate congruence using Holland codes three letters in length and that congruence should only be calculated using those Holland types that are clearly salient in the profiles being compared. The modifications to the C index proposed in this article allow comparisons between Holland code profiles of unequal lengths and/or of less than three letters in length and retain the desirable properties of the original C index: reflection of the hexagonal structure, sensitivity to order, and simplicity of calculation.

Keywords: C index, Holland types, Holland codes, RIASEC, person-environment fit, congruence

Holland's (1997) theory of career types has enjoyed widespread acceptance for several decades and has spurred the generation of many hundreds if not thousands of publications. The incorporation of Holland's model into the Occupational Information Network (O*NET; U.S. Department of Labor, 1998) should ensure its continued use for many years to come.

In simplest terms, Holland's (1997) theory proposes that individuals and occupations may be described in terms of six primary personality types: realistic, inves-

Correspondence concerning this article should be addressed to Donald E. Eggerth, Training Research and Evaluation Branch, CDC/NIOSH, 4676 Columbia Parkway, C-10, Cincinnati, OH 45226; e-mail: deggerth@cdc.gov.

JOURNAL OF CAREER ASSESSMENT, Vol. 14 No. 2, May 2006 267–275
DOI: 10.1177/1069072705283976

tigative, artistic, social, enterprising, and conventional. A profile containing combinations of Holland types is used to describe the salient characteristics of an individual or of a given occupation. These combinations are referred to as *Holland codes*. As a shorthand convention, the Holland types are frequently referred to by using only the first letter of each type's full name, hence the terms *code letters* and *RIASEC model*.

At the core of Holland's (1997) theory is the proposition that the more similar, or congruent, the Holland codes for an individual and for an occupation are, the better the fit between worker and job. Individuals who are in occupations with congruent Holland codes are expected to be psychologically healthier, more satisfied, and more productive than individuals in incongruent occupations.

ASSESSING CONGRUENCE

In his most recent statement of his theory of career types, Holland (1997) endorsed the C index (Brown & Gore, 1994) as the best Holland code congruence index yet developed. In his critical review of the RIASEC person-environment (P-E) fit literature, Tinsley (2000) identified the C index as one of the best Holland code congruence indices because it operationalized the hexagonal structure of the RIASEC model and was sensitive to code order. Moreover, the C index was far simpler to calculate than most other Holland code congruence indices.

Similar to nearly all other commonly used Holland code congruence indices, the C index is calculated using the three highest ranking Holland codes for the profiles being compared (Brown & Gore, 1994). This formulation works well when using the instrumentation and supporting materials developed by Holland and his colleagues. For example, the Holland codes contained in both the Self-Directed Search (SDS) Occupations Finder (Holland, 1996) and the *Dictionary of Holland Occupational Codes* (DHOC; Gottfredson & Holland, 1996) are always three code letters in length. However, other major sources of Holland codes do not follow this practice. Both the Strong Interest Inventory (SII; Harmon, Hansen, Borgen, & Hammer, 1994) and the Occupational Information Network (O*NET; Rounds, Smith, Hubert, Lewis, & Rivkin, 1999) use decision rules to determine whether a particular Holland type is salient enough to be included in the code used to characterize a given occupation. The Holland codes used by the SII and the O*NET vary in length from one to three letters. Consequently, researchers wishing to use the C index will experience difficulty working with Holland code sources other than the SDS Occupations Finder and the DHOC.

Gati (1985) criticized the practice of always characterizing individuals and occupations using three code letters and argued that comparisons between Holland code profiles should only be made using those Holland types that are most salient in each profile. Gati developed the computationally complex Sb index to allow for comparisons between profiles shorter than three letters and/or of unequal lengths. However, Gati's aspect-based approach to P-E fit does not consider the relative ordering of salient Holland types to be significant, and the Sb index was not formulated to be sensitive to code order. For example, the Sb index would consider the Holland code RIA to be equally well matched by the codes RIA, RAI, IRA, IAR, AIR, or ARI. Although this interpretation is consistent with the instructions for the SDS Occupations Finder (Holland, 1996) that encourage users to explore all possible combinations of their Holland codes, it is not consistent with the interpretation and application of the Holland RIASEC model by the majority of vocational researchers. Therefore, despite its ability to make comparisons between Holland codes of unequal lengths, the usefulness of the Sb index may be limited for some researchers.

In work comparing the Holland code classification agreement rates across the DHOC, the SII, and the O*NET, Eggerth, Bowles, Tunick, and Andrew (2005) called for the development of a Holland code congruence index that would combine the sensitivity to order of the C index with the ability of the Sb index to make comparisons between profiles shorter than three letters and/or of unequal length. This article proposes a modification of the C index that has these desirable properties.

CALCULATING CONGRUENCE

In general terms, the calculation of the C index, like that of most Holland code congruence indices, can be broken down into three steps.

1. Quantifying the hexagonal distance between the Holland types being compared
2. Weighting the values determined in Step 1 depending on whether the Holland types being compared occupied the first, second, or third letter position
3. Summing the weighted values from Step 2

THE C INDEX

Specifically, given two Holland code profiles $X_1X_2X_3$ and $Y_1Y_2Y_3$, with the subscripts representing the first, second, and third code letters as indicated, the C index may be expressed as follows:

$$C = 3\,(X_1,Y_1) + 2\,(X_2,Y_2) + 1\,(X_3,Y_3)$$

where (X_i,Y_j) are the values assigned to each comparison between Holland code profiles on the basis of the hexagonal distance between the letters X_i and Y_j. An identical match is assigned a value of 3. An adjacent position is assigned a value of 2. A nonadjacent position is assigned a value of 1. An opposite hexagonal position is given a value of 0. The first, second, and third code letter comparisons are weighted 3, 2, and 1, respectively. The value of the C index can range from 0 when comparing profiles that are hexagonal opposites to 18 when comparing identical profiles.

MODIFYING THE C INDEX

The rationale for the proposed modifications to the C index is based on an acceptance of Gati's (1985) argument that Holland code profile comparisons should only be made using the most salient Holland types in a profile. It is our argument that some profiles may be fully characterized using only one Holland type, whereas others may require two or three code letters to capture all that is meaningful about the profile. This is essentially the principle that the developers of both the SII and the O*NET have used to guide their Holland code assignments.

Using the notation introduced above, a profile requiring three Holland types to adequately describe it would be $X_1X_2X_3$. A profile requiring only two Holland types would be X_1X_2, and one requiring a single Holland type would be X_1. Given that rational decision rules are used to assign only salient Holland types to a code profile, it is our argument that each Holland code profile, regardless of length, contains the same quantity of descriptive information. Therefore, when one is comparing profiles of unequal length, the shorter profile merely represents a greater "concentration" of information than does the longer profile. In shorter profiles, the Holland types that are included carry all the information of the "missing" types. This concept is perhaps best understood through several examples.

Let us consider a case in which one wishes to compare a Holland code profile that needs three different types to describe it with another profile that can be adequately described using only one type. Using our notation, these profiles would be $X_1X_2X_3$ and Y_1. It is our argument that this comparison should be viewed as the equivalent of comparing the profiles $X_1X_2X_3$ and $Y_1Y_1Y_1$. That is to say, if only Y_1 is needed to characterize all that is salient about a profile, it carries all the information of the "missing" two types (Y_2 and Y_3).

Let us now consider a case in which one wishes to compare two Holland code profiles of two letters in length (X_1X_2 and Y_1Y_2). This comparison might be conceptualized as comparing the profiles $X_1X_2WC_x$ and $Y_1Y_2WC_y$. In this case the "missing" third letters from each profile are each represented by a weighted composite (WC_x and WC_y) of the first two letters of their respective profiles (X_1X_2 and Y_1Y_2). Following the logic common to most Holland code congruence indices, the composite is weighted so that the first code letter contributes more to the overall estimate of congruence than does the second code letter.

The specific formula for calculating the contribution of this weighted composite to the overall estimate of congruence is presented later in this article. However, in all cases but one, mathematically equivalent results can be achieved with simpler formulas that use a scaling constant. For the sake of clarity, we have presented these simpler computational formulas in the following discussion. This presentation will also make it clearer that these are modifications to the C index and do not represent an entirely new congruence index.

SIX POSSIBLE CASES

When one is comparing two Holland code profiles, each of which may be up to three letters in length, there are six possible combinations of code pairings. Let us consider each pairing in turn. Please note that in all of the following equations, the definitions remain the same as used in the description of the C index above. We will begin with comparisons of profiles of equal length.

Case 1: Three Letters × Three Letters

This combination requires no modification of the current C index. Case 1 index may be expressed as

$$C_{3\times3} = 3\ (X_1,Y_1) + 2\ (X_2,Y_2) + 1\ (X_3,Y_3).$$

We have added the subscript 3 × 3 to indicate that this is the formula for calculating congruence when comparing profiles that are each three letters in length. We follow the same convention for the remaining computational formulas to clarify the instance of appropriate application. As discussed earlier, the $C_{3\times3}$ index ranges from 0 when comparing profiles that are hexagonal opposites to 18 when comparing identical profiles.

Case 2: Two Letters × Two Letters

The starting point for calculating congruence between profiles of this combination is the first two terms of the C index. This may be expressed as

$$C_{\text{Case 2}} = 3\ (X_1,Y_1) + 2\ (X_2,Y_2).$$

As written, the index would range from 0 when comparing profiles that are hexagonal opposites to 15 when comparing identical profiles. In Case 1, identical profiles have a value of 18. However, if these two-letter Holland codes contain the same amount of descriptive information as the three-letter codes in Case 1, it makes no sense that a perfect fit between two-letter profiles should count for less (15) than a perfect fit between three-letter profiles (18). This may be easily remedied by using a multiplicative constant to put the $C_{2\times2}$ index on the same metric as the $C_{3\times3}$ index. Use of a scaling constant with the value of 18/15 will do so. Therefore, the final form of Case 2 is

$$C_{2\times2} = 18/15\ [3\ (X_1,Y_1) + 2\ (X_2,Y_2)].$$

Case 3: One Letter × One Letter

The starting point for calculating congruence between profiles of this combination is the first term of the C index. In this case, the single-letter comparison would have a range of 0 to 9. Following the same logic as presented for Case 2, a scaling constant of 18/9 or 2 is used to give the Case 3 index the same range as the C index. The final form of Case 3 is

$$C_{1\times1} = 2\ [3\ (X_1,Y_1)],$$

which reduces further to

$$C_{1\times1} = 6\ (X_1,Y_1).$$

With this combination, some might suggest that one should use the familiar first-letter hexagonal distance index (Holland, 1973). In a sense, given that Brown and Gore (1994) considered the C index to be a simple extension of the first-letter hexagonal distance index, that is exactly what we are proposing. However, Brown and Gore changed the hexagonal distance values used by the first-letter hexagonal distance index from 4, 3, 2, 1 to 3, 2, 1, 0 for the C index. Therefore, the Case 3 index might be thought of as merely a modification of the familiar first-letter hexagonal distance index that uses slightly different hexagonal distance values and a scaling constant to keep its values in the same range as the $C_{3\times3}$ index.

We will now discuss comparison of profiles of unequal length.

Case 4: Three Letters × One Letter

This is the same case that was used earlier as an example. Using our notation, these profiles would be written as $X_1X_2X_3$ and Y_1. For the purpose of congruence calculations, it is our argument that this is the equivalent of comparing the profiles $X_1X_2X_3$ and $Y_1Y_1Y_1$. One may calculate congruence between these two profiles by using the C index as currently formulated, with the exception that in calculating the hexagonal distances (X_i,Y_j), Y_j only takes on the value of Y_1. This may be expressed as

$$C_{3\times1} = 3\ (X_1,Y_1) + 2\ (X_2,Y_1) + 1\ (X_3,Y_1).$$

Case 5: Two Letters × One Letter

The Case 5 modification of the C index combines the logic of the Case 2 and the Case 4 modifications. Following Case 4, this is viewed as comparison of the profiles X_1X_2 and Y_1Y_1. Following Case 2, a scaling constant of 18/15 is used to give Case 5 the same possible range as the unmodified C index. Case 5 may be expressed as

$$C_{2\times1} = 18/15\ [3\ (X_1,Y_1) + 2\ (X_2,Y_1)].$$

Case 6: Three Letters × Two Letters

Case 6 represents the most complex of our proposed modifications to the C index because it is the instance in which calculations involving a weighted composite cannot be avoided by using a scaling constant. This case may be conceptualized as representing a comparison between the profiles $X_1X_2X_3$ and $Y_1Y_2WC_y$, where the term WC_y represents a weighted composite of Y_1 and Y_2. The rationale behind using a composite of Y_1 and Y_2 is essentially the same as that used in Cases 4 and 5, in which Y_1 was used in place of Y_2 and Y_3. However, when only a single type is used to "replace" the "missing" types, no attention need be paid to ordering, and therefore no weighting is necessary.

When two types are used as a composite to replace a "missing" third-letter type, the logic behind the original formulation of the C index demands that they be differentially weighted. Recall that the C index uses the weights 3, 2, and 1 for the first, second, and third code letter positions, respectively. By this logic, when using the first two types to form a composite to "replace" the missing third code letter, the ratio of the weighting of the first type to the second type should be 3:2.

The formula for calculating the contribution of this weighted composite to the overall estimate of congruence may be represented as

$$WC = 1/5\ [3\ (X_3,Y_1) + 2\ (X_3,Y_2)].$$

Substituting this term into the calculation for the full index yields,

$$C_{3\times2} = 3\ (X_1,Y_1) + 2\ (X_2,Y_2) + 1\ \{1/5\ [3\ (X_3,Y_1) + 2\ (X_3,Y_2)]\}.$$

SUMMARY

Following Holland's lead with the SDS (Holland, 1994, 1996) and the DHOC (Gottfredson & Holland, 1996), most researchers characterize both individuals and occupations using their three highest Holland types. In most cases, this assignment is made solely on the basis of score rankings without consideration of score elevation or differentiation.

We agree with Gati (1985) that this method of assigning Holland types to profiles risks being arbitrary and that only those Holland types that are salient should be used to characterize a Holland code profile. The Holland codes assigned to occupations by the SII and the O*NET vary from one to three letters in length. Unfortunately, the C index, which is arguably the best of the existing Holland code congruence indices, is only formulated to make comparisons between profiles three code letters in length. As a result, the C index cannot easily be used to assess congruence using the Holland codes for occupations from either the SII or the O*NET.

The modifications to the C index proposed in this article will allow it to be used to make comparisons between Holland code profiles of unequal lengths and/or of less than three letters in length while retaining the desirable properties of reflecting the theorized hexagonal structure of the Holland RIASEC model and sensitivity to code letter ordering. Simplicity of calculation is maintained in all cases but one by the use of scaling constants to maintain a common metric. Even in the one case where a scaling constant cannot be used (the $C_{3\times2}$ index), its calculation is considerably simpler than that of the Sb index, the only other congruence measure that allows comparison of Holland code profiles of unequal lengths.

A Final Thought

We recognize that the methods proposed in this article beg the question of how elevated a Holland type needs to be in order to be considered salient enough to be included in the Holland code profile for either an individual or an occupation. Although this is clearly an important topic, it is beyond the intended scope of this article to discuss it in the detail that it deserves. Both the SII and the O*NET use decision rules to determine whether a particular Holland type is salient enough to merit being included in the code used to characterize a given occupation. Perhaps these rules could serve as useful starting points for assigning Holland types to the profiles of individuals. Ultimately, the best method for assigning Holland types to a profile remains to be determined.

REFERENCES

Brown, S. D., & Gore, P. A. (1994). An evaluation of interest congruence indices: Distribution characteristics and measurement properties. *Journal of Vocational Behavior*, *45*, 310-327.

Eggerth, D. E., Bowles, S. M., Tunick, R. H., & Andrew, M. E. (2005). Convergent validity of O*NET Holland code classifications. *Journal of Career Assessment*, *13*, 150-168.

Gati, I. (1985). Description of alternative measures of the concepts of vocational interest: Crystallization, congruence, and coherence. *Journal of Vocational Behavior*, *27*, 37-55.

Gottfredson, G. D., & Holland, J. L. (1996). *Dictionary of Holland occupational codes* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.

Harmon, L. W., Hansen, J. C., Borgen, F. H., & Hammer, A. L. (1994). *Strong Interest Inventory applications and technical guide*. Palo Alto, CA: Consulting Psychologists Press.

Holland, J. L. (1973). *Making vocational choices: A theory of careers*. Englewood Cliffs, NJ: Prentice Hall.

Holland, J. L. (1994). *Self-directed search*. Lutz, FL: Psychological Assessment Resources.

Holland, J. L. (1996). *The occupations finder*. Odessa, FL: Psychological Assessment Resources.

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.). Odessa, FL: Psychological Assessment Resources.

Rounds, J., Smith, T., Hubert, L., Lewis, P., & Rivkin, D. (1999). *Development of occupational interest profiles for* O*NET. Raleigh, NC: National Center for O*NET Development.
Tinsley, H. E. A. (2000). The congruence myth: An analysis of the efficacy of the person-environment fit model. *Journal of Vocational Behavior*, *56*, 147-179.
U.S. Department of Labor. (1998). *O*NET 98 viewer user's guide for version 1.0*. Washington, DC: Government Printing Office.