

Article

Concepts and Coefficients Based on John L. Holland's Theory of Vocational Choice—Examining the R Package *holland*

Florian G. Hartmann ^{1,*}, Jörg-Henrik Heine ^{2,†} and Bernhard Ertl ³

¹ Department of Human Sciences, Institute for Education, Methodology in the Social Sciences, Universität der Bundeswehr München, 85577 Neubiberg, Germany

² Center for International Student Assessment (ZIB), TUM School of Social Sciences and Technology, Department of Educational Sciences, Technical University Munich, 80333 München, Germany; joerg.heine@tum.de

³ Department of Human Sciences, Institute for Education, Learning and Teaching with Media, Universität der Bundeswehr München, 85577 Neubiberg, Germany; bernhard.ertl@unibw.de

* Correspondence: florian.hartmann@unibw.de

† These authors contributed equally to this work.

Abstract: John L. Holland's theory of vocational choice is one of the most prominent career theories and is used by both researchers and practitioners around the world. The theory states that people should seek work environments that fit their vocational interests in order to be satisfied and successful. Its application in research and practice requires the determination of coefficients, which quantify its core concepts such as person-environment fit. The recently released R package *holland* aims at providing a holistic collection of the references, descriptions and calculations of the most important coefficients. The current paper presents the package and examines it in terms of its application for research and practice. For this purpose, the functions of the package are applied and discussed. Furthermore, recommendations are made in the case of multiple coefficients for the same theoretical concept and features that future releases should include are discussed. The R package *holland* is a promising computational environment providing multiple coefficients for Holland's most important theoretical concepts.

Keywords: vocational choice; Holland; RIASEC; congruence; differentiation; consistency



Citation: Hartmann, F. G.; Heine, J.-H.; Ertl, B. Concepts and Coefficients Based on John L. Holland's Theory of Vocational Choice—Examining the R Package *holland*. *Psych* **2021**, *3*, 728–750. <https://doi.org/10.3390/psych3040047>

Academic Editor: Holmes Finch

Received: 30 August 2021

Accepted: 23 November 2021

Published: 29 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

John L. Holland's [1] theory of vocational choice is a major career theory [2] and as such is one of the most well known and widespread approaches in vocational psychology [3]. The theory has been extensively used around the world by both researchers and practitioners. From 1953 to 2016, it has been cited by over 2300 works in over 350 different journals and periodicals [4]. Holland's theory found its way into practice especially by implementing career consulting tools such as the Self-Directed Search ([SDS]; [5]) that helps people find a work environment that fits their vocational interests. The SDS has been translated into 25 languages and is estimated to be used by more than two million people each year [6].

In addition, the SDS provided a basis not only for its mere translation, but also for new and further developments of inventories for assessing vocational interest orientations. For German speaking countries for example the *Allgemeiner Interessen Struktur Test* ([AIST] [7–9]) offers a good implementation for diagnosing vocational interests in research, as well as applied settings. For diagnosing vocational interests in the applied setting of secondary education in German language areas, the EXPLORIX [10] can be used.

Applying Holland's theory, whether it is to explain vocational behavior or to help make promising career choices, requires the determination of coefficients that describe individuals and their (potential) work environments. This can be a complex and time-consuming process, especially for researchers or career counselors that apply Holland's

theory for the first time. Since the same concept (e.g., person-environment fit) can be quantified in numerous different ways, the mere selection of a coefficient that suits the research or counseling goal can be a challenge. With the recently released R package *holland* [11] we try to deal with this problem by providing a sound basis of core functionalities including the calculation of the most important coefficients in the framework of Holland's theory.

2. The Present Paper

Given the initial situation briefly outlined above, the present paper aims at examining the applications the R package *holland* offers for research and practice based on Holland's theory. We first provide an introduction into Holland's theory to contextualize the basic functionalities of the package for both—research settings in the frame of vocational psychology as well as practical, applied settings around vocational counseling. Subsequently, in an applied section, we then aim at outlining the structure and functions of the R package *holland*. By means of short R snippets we demonstrate how the concepts of Holland's [1] theory are implemented and what coefficients regarding their operationalization are provided. In the case of multiple coefficients for the same concept we aim at giving recommendations depending on the context and goal of application. Finally, we discuss which functions should be provided by future releases of the package.

3. Holland's Theory of Vocational Choice

Holland's [1] theory is a person-environment fit approach stating that people seek work environments that suit their interests and that the degree of fit between a person and a specific work environment has substantial effects on favorable outcomes such as job satisfaction and performance [3]. To describe Holland's theory in more detail, it is useful to distinguish between its *central assumptions* and its *concepts* (individual constructs) [12]. Moreover, in this chapter, different methods to describe persons and work environments regarding their interest profiles are explained since these methods have direct effects on which coefficients can finally be determined in order to quantify the concepts of Holland's theory (for a more comprehensive description see also [13]).

3.1. Central Assumptions

Holland [1] claims that individuals and their work-related preferences can best be described using six broad interest dimensions. These dimensions are Realistic (R), Investigative (I), Artistic (A), Social (S), Enterprising (E), and Conventional (C), which is why Holland's theory is often referred to as the RIASEC model. Every dimension comprises the interest in specific activities. The Realistic dimension corresponds to the preference for producing concrete and visible things by using tools and machines, the Investigative dimension to the preference for dealing with complex problems by means of science, the Artistic dimension to the fondness for unstructured activities that require creativity, the Social dimension relates to teaching, helping or caring, the Enterprising dimension to persuading and leading other people, and the Conventional dimension finally corresponds to the interest in the orderly and systematic handling of data. One of Holland's most important assumptions is that not only people but also their (potential) work environments can be characterized using the very same six RIASEC dimensions. For example, a Realistic occupation would be one that requires Realistic interests, i.e., the preference for working with tools and machines, and would reward those inhabitants who show just those interests.

According to Holland's *calculus* hypothesis the six interest dimensions are not orthogonal to each other but can be arranged as a regular hexagon in a two dimensional space with the spatial distance between the RIASEC dimensions reflecting their psychological similarity. For example, the Investigative dimension is more similar to the Artistic dimension than to the Social dimension and contradictory to the Enterprising dimension. This also means that it is rather unlikely that a person shows both Investigative and Enterprising interests to a strong extent. As Figure 1 illustrates, a regular hexagon lies on a circle, which is why

the hexagonal arrangement of the RIASEC dimensions constitutes a circumplex structure of vocational interests (e.g., [14]). A structural meta-analysis confirmed the calculus hypothesis with regard to people and work environments [15]. To test whether RIASEC data fits the hexagonal model, manifest variable approaches and latent variable approaches can be distinguished. A prominent example for a manifest variable test is the *randomization test of hypothesized order relations* ([RTOHOR]; [16]). Tracey wrote a software program [17] as well as an R code that both provide the RTOHOR and that both can be obtained on his homepage: <http://tracey.faculty.asu.edu/index.html> (accessed on 20 November 2021).

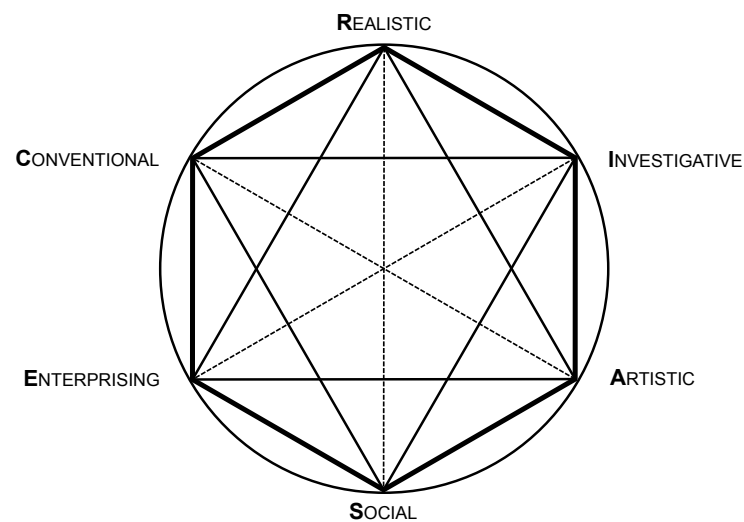


Figure 1. Hexagonal arrangement of the RIASEC dimensions.

Figure 2 shows the illustration of a latent variable approach [18], where the estimated angular locations of the RIASEC dimensions are projected into the perfect hexagon. Here, it is also possible to project additional variables, such as the Big Five which have been found to be related to the RIASEC dimensions [19].

Another central assumption of the theory is that people seek work environments that fit their interests and allow them to act according to their personal preferences. For example, a person that has Enterprising interests in the first place is attracted to Enterprising work environments and vice versa. This means that an Enterprising environment seeks for people who are interested in leading other people and whose preferences can best be described by the Enterprising interest dimension.

The vocational behavior of an individual now results from the interaction between that person and her or his environment. For example, if a person's interests do not fit the current job, she or he is supposed to leave and search for another better fitting work environment [20].

3.2. Concepts

Holland [1] defines several concepts that go beyond the information of the RIASEC dimensions and characterize a person or a work environment more comprehensively; some of these concepts play a more important role in research and practice than others. Because this paper examines an R package that aims at covering Holland's theory, we will briefly outline *all* of these concepts.

The most important concept (or individual construct) within Holland's theory, which may also be a main goal of career counseling, is called *congruence* and is given if a person chooses a work environment that fits her or his vocational interests. According to Holland [1], congruent choices are supposed to be related to job satisfaction, persistence and performance. This congruence hypothesis was confirmed by several meta-analyses [21–23].

The concept of *consistency* is based on the calculus hypothesis. If the circumplex structure appears in an individual interest profile of a person, she or he is consistent according to Holland's theory. For example, a consistent person with strong Realistic interests could also show strong Investigative or Conventional interests but should score low on the Social dimension (see Figure 1).

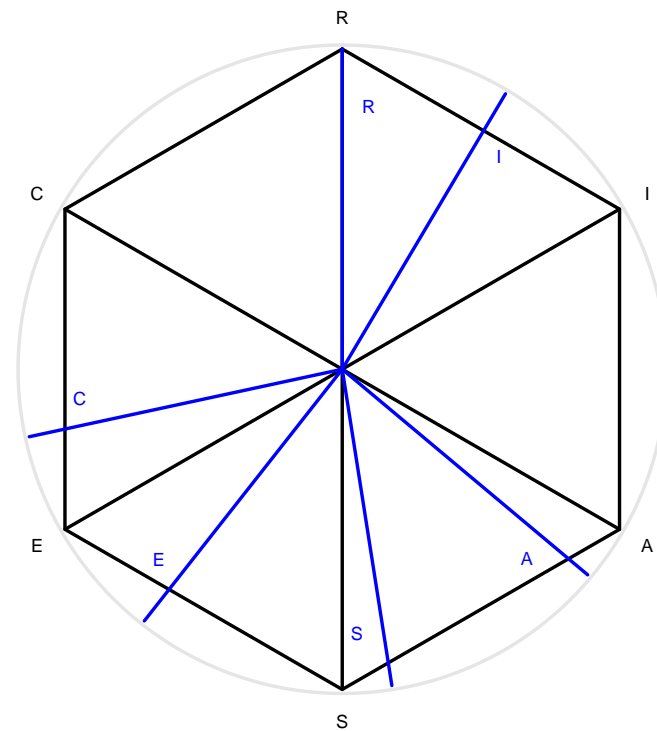


Figure 2. Projection of the RIASEC dimensions into the perfect hexagon based on a sample of a South German University ($N = 734$) [24].

Differentiation is another concept and is given if a person scores high on only one (or few) RIASEC dimension(s) and scores low on the remaining dimensions. A person who shows equally strong RIASEC interests would be little differentiated.

Consistency and differentiation are supposed to predict congruence and desirable outcomes such as persistence and performance since consistent individuals do not have to reconcile conflicting interests and differentiated individuals have a clear interest profile that eases the search for a congruent work environment [1,12,25]. Evidence regarding the effects of consistency and differentiation is somewhat equivocal, which may also be due to more or less sophisticated measurement methods that have been applied [25].

Another concept is *identity* and can be defined as clarity a person has regarding her or his talents, interests and vocational goals [1]. While the aforementioned constructs can be derived from an interest profile, i.e., the RIASEC scores of a person, measuring identity requires additional assessment tools see (e.g., [26]).

Initially *elevation* has not been defined as a central concept of Holland's theory but has grown in importance since it is seen as the interest flexibility a person exhibits [27]. It is simply measured as the mean or sum of a person's RIASEC scores and indicates the personal interest level.

As with the RIASEC dimensions, the concepts of consistency, differentiation, identity, and elevation can also be applied to work environments in order to characterize them in more detail.

3.3. Characterizing People and Work Environments Using the RIASEC Dimensions

The methods used to characterize people and environments with the RIASEC dimensions have an impact on which coefficients can be determined in order to operationalize the central concepts (e.g., congruence) of Holland's [1] theory.

Traditionally, an individual or environmental interest profile has been represented by a so-called *three-letter code*, which consists of the first letters of the three strongest interest dimensions. For example, if a person shows strong Social interests, followed by Investigative and Artistic interests, this person would be assigned the three-letter code *SIA*. The same method applies for work environments [1]. If a work environment requires Social interests in the first place followed by Investigative and Conventional interests, this environment would be assigned the three-letter code *SIC*. In some studies, only the strongest or the two strongest RIASEC dimensions are used to describe a person or an environment (e.g., [28]). These *high-point codes* are obviously associated with a loss of information, since the weaker of the six RIASEC dimensions are not taken into account. That is why scholars suggest to use *full profile information*, i.e., all six RIASEC scores, to comprehensively describe a person or a work environment [29,30].

To generate a person's RIASEC profile, a vocational interest inventory can be used, which usually comprises of (at least) six scales delivering six RIASEC scores (e.g., [5,9,31,32]). If required, a one-, two- or three-letter high-point code can be derived by ranking the RIASEC scores.

Generating RIASEC scores (or high-point codes) for a work environment is more challenging. Here, different methods can be distinguished [33], with the *incumbent method* or the *judgment method* being usually applied. The incumbent method is based on the idea, that people determine the environments they are working in [1]. Correspondingly, researchers applying the incumbent method use and aggregate the RIASEC interest scores of the people of a specific work environment in order to generate the respective environmental RIASEC scores or high-point code. Since the incumbent method needs a lot of resources, i.e., representative samples for every work environment [33,34], it is especially useful if only one or few different work environments are examined, so that the respondents of the study can be seen as the incumbents and their RIASEC scores can be used to create the environmental RIASEC profiles. In comparison to the incumbent method, the judgment method requires far less resources. It uses the ratings of trained judges due to descriptions of work environments to create environmental RIASEC profiles. The judgment method was used to create the occupational RIASEC profiles of the Occupational Information Network ([O*NET]; [35]), a system that was developed by the U.S. Department of Labor's (USDOL's) Office of Policy and Research and provides, among others, extensive information regarding occupational characteristics. If required, the O*NET RIASEC profiles [33,36] can be assigned not only to US occupations with a O*NET-SOC code but via a crosswalk also to nationally coded occupations of other countries (see e.g., [37,38]; please note: the O*NET-SOC taxonomy is an occupational classification with codes assigned to the different classes. It is based on the US Standard Occupational Classification (SOC), which "is used by federal statistical agencies to classify workers and jobs into occupational categories for the purpose of collecting, calculating, analyzing, or disseminating data" ([39], p. 1). In its current version, the O*NET-SOC taxonomy comprises of 1016 occupational titles. In an investigation, the occupations indicated by respondents in the US can be directly assigned O*NET-SOC codes and as a result, are linked to the RIASEC interest profiles of the O*NET. Please see the O*NET Resource Center for more information: <https://www.onetcenter.org/taxonomy.html> accessed on 20 November 2021). This is of special interest for countries in which no RIASEC scores but only high-point codes are available with respect to occupations, since the state-of-the-art coefficients that are used for the quantification of Holland's concepts usually require full RIASEC profile information.

In addition to high-point codes and full RIASEC score profiles, Eder [40] proposed another method to describe persons and work environments in the framework of Holland's theory. Here, the RIASEC scores of a personal or an environmental profile are transferred into vectors in a two dimensional space under consideration of the calculus hypothesis. The sum of these six RIASEC vectors lead to an *interest vector* that points in one of the RIASEC dimensions and considers full profile information as well as the assumed hexagonal structure of the RIASEC dimensions. The coordinates of such a vector can also be determined by calculating the scores on Prediger's [41] people/things (P/T) and ideas/data (I/D) dimensions.

3.4. Coefficients Based on the RIASEC Model

The operationalization of the concepts described above implies the computation of coefficients, many of which were developed specifically for Holland's [1] theory. Which coefficients can be determined also depends on whether RIASEC scores or high-point codes are used to characterize people and work environments (see Table 1). To be able to give recommendations for use, in the application part of the paper, we will exemplify and compare the various coefficients and algorithms.

Regarding the operationalization of *congruence* Xu and Li [42] distinguish between *top-letter(s)* methods, *profile-based empirical* methods and *profile-based conceptual* methods. Using the top-letter(s) approach, based on the comparison of the personal high-point code and the environmental high-point code, coefficients, so-called *congruence indices*, are calculated to quantify congruence. For example, a very simple measure, the *dichotomous first-letter agreement index* [28], compares the dominant personal RIASEC dimension with the dominant environmental RIASEC dimension, and is 1 if the dimensions are identical (e.g., R and R) and 0 if not (e.g., R and I). A large number of different congruence indices have been developed [43,44]. To give a rough overview, these indices can be classified according to, among other things, whether they take one ([e.g., *first-letter agreement based on the hexagon*]; [45]), two ([e.g., *two-letter agreement index*]; [46]) or three ([e.g., *M3 index*]; [47]) RIASEC dimensions into account, whether they consider the hexagonal structure of interests ([e.g., *C index*]; [43]), whether they allow asymmetric comparisons between two high-point codes with different numbers of letters ([e.g., *Modified C index*]; [48,49]), and whether they weight the RIASEC dimensions within a high-point code ([e.g., *M index*]; [50]) or not ([e.g., *N-3 index*]; [51]). If the profile-based empirical approach is used, congruence is indicated by congruence outcome scores (e.g., performance scores) that are estimated based on personal and environmental RIASEC profiles ([e.g., *polynomial regression*]; [52]). Within the profile-based conceptual approach, at least four congruence measures can be distinguished [42]: profile correlation, profile deviance, Euclidean distance, and angular agreement. All these measures require six personal and six environmental RIASEC scores. The profile correlation is simply the Pearson correlation between the personal and the environmental RIASEC profile [3]. The profile deviance is determined by aggregating the squared differences between the personal and environmental RIASEC scores and subsequently taking the square root of this expression [3]. In contrast to profile deviance and profile correlation, the Euclidean distance and the angular agreement are based on the calculus hypothesis and are using the personal and environmental RIASEC vectors in the two dimensional space to quantify congruence. The Euclidean distance indicates the distance between the two vector peaks, so that higher values indicate lower congruence [53]. The angular agreement is based on the A index [54] and the angle between the personal and the environmental vector so that higher values indicate higher congruence [53].

Xu and Li [42] review the literature regarding the criterion validity of different congruence measures and suggest to preferably use the profile-based conceptual approach. Based on a comparative examination, they also declare the profile correlation as a superior coefficient to quantify Holland's concept of congruence.

Table 1. Concepts and corresponding coefficients based on Holland's [1] RIASEC dimensions.

Concept	Profiles	Coefficient
Congruence	1-letter codes	DFLA index [28]
		FLAH [45]
		TLA index [46]
	2-letter codes	Z-S index [55]
		TLC index [56]
	3-letter codes	Compatibility index [57]
		RCCS [58]
	Flexible number of letters	M index [50]
		K-P index [59]
		C index [43]
		M3 index [47]
		N-3 index [51]
		Sb index [60]
Modified C index [48,49]		
Hamming distance [61]		
Levenshtein distance [62]		
PICS [63]		
Scores and 3-letter codes	Polynomial regression [52]	
	Profile correlation [3]	
	Profile deviance [3]	
	Euclidean distance [53]	
	Angular agreement [53]	
Consistency	2-letter codes	Holland's measure [64]
		Strahan's measure I [65]
	3-letter codes	Strahan's measure II [65]
		Scores
Differentiation	Scores	Frantz & Walsh's measure [66]
		Spokane & Walsh's measure [67]
		Iachan's measure I [68]
		Iachan's measure II [68]
		Holland's measure [45]
		Peiser & Meir's measure [69]
		Healy & Mourton's measure [46]
		Length of interest vector [40]
Amplitude (cosine function) [25]		
Interest flexibility	Scores	Sum of scale scores [27]

Notes. Abbreviations by the authors: DFLA index = Dichotomous first-letter agreement index; FLAH = First-letter agreement based on the hexagon; TLA index = Two-letter agreement index; TLC index = Three-level congruence index. Coefficients that are provided by the R package *holland* are in bold. See the discussion section for an elaboration on the sufficiency of the R package in its current state.

With regard to *consistency* there are again measures that use only a few of the six RIASEC dimensions. For example, Strahan [65] developed two measures one of which uses two-letter codes and one of which uses three-letter codes. Holland also proposed a measure that compares only the first and second letter of a RIASEC high-point code based on the calculus hypothesis [64].

A traditional measure for the *differentiation* of a personal or an environmental RIASEC profile is the difference between the highest and the lowest RIASEC score [45]. In addition, similar measures have been developed that do not take full profile information into account (e.g., [66–68]). In contrast, Healy and Mourton [46] suggest to use a full RIASEC score profile and to determine its standard deviation as an indicator for differentiation. Another measure for differentiation that is based on the calculus hypothesis is the length

of the aforementioned interest vector [40,41]. Based on the cosine function, the measures proposed by Tracey et al. [25] also use full RIASEC score profile information and refer to the circumplex structure of vocational interests to operationalize differentiation and also consistency.

Interest flexibility can simply be measured by determining the sum of the individual RIASEC scores [27].

4. The R Package *holland*

The R package *holland* is hosted on CRAN since the 22nd of June 2021 (see [11]). Its provision on CRAN is based on an initially loose R function collection on Holland's theory, which, beginning in 2011, was expanded increasingly in the Working Group for Social Science Methods at the Universität der Bundeswehr München. The package offers a convenient way to compute parameters in the framework of Holland's theory [1]. As a start, at least the six RIASEC interest scores are required for each person so that the concepts of Holland's theory can be determined on an individual level using the R package *holland* (please see the data set in Section 4.4 for an example). In its current version, the package provides three main functional areas.

One functional area is related to the concept of *congruence* between a person's interest orientation and a particular vocational environment. For this, the package (currently) offers ten R-functions with which different indices for the congruence can be calculated (see all functions starting with 'con_', (e.g., `con_oneletter_holland()`). The second function area is related to the concept of *differentiation*, which is currently only covered with the function `dif_7_holland()` to compute seven different indices for differentiation. The last main function area addresses the so-called *calculus* hypothesis, according to which the six interest orientations are arranged in the form of a hexagonal structure. The package *holland* offers, among other functions, three (wrapper) functions, which are directly addressed to the user. Within the calculus hypothesis the arrangement of empirical data can be determined (cf. function `Circ_emp`) and their fit to the hexagonal structure can be determined (cf. function `Circ_test`). Furthermore, other construct domains (e.g., Big Five personality) with their dimensions can be projected into the hexagonal structure (cf. Function `Circ_pro`). These three functions are based on the method of structural equation modeling proposed by Nagy et al. [18], which was implemented as Mplus syntax. The application of the three functions therefore requires an installation of the commercial software Mplus (cf. also `MplusAutomation`). In addition to these three core areas, there is also the *misc* area, which includes various auxiliary functions, and the *datasets* area, which includes various correlation tables for the RIASEC (captured with the AIST-R [8,9]) and Big Five dimensions.

In the following sections, the three core areas and the area miscellaneous are briefly represented by means of short R code snippets. In order to get familiar with the structure of *holland* and run the example R-code snippets, the latest version of *holland* from the CRAN repositories has to be installed. This is done for users of R-Studio either via the menu control in R-Studio (packages → install → search term: 'holland') or by simply entering the following R command into the console (R-Studio is a convenient R editing and development environment, which offers some convenient tools for managing R packages, multiple R workspaces and R-code scripts. It is available for the three most popular operating systems: Linux, Mac OS and Windows; <https://www.rstudio.com/>, accessed on 20 November 2021).

```
install.packages("holland",dependencies = TRUE)
```

4.1. Congruence Indices in *holland*

For the area of *congruence*, there are currently 10 separate functions to compute different congruence indices. All 10 functions calculate the extent of congruence between two sequences of high-point codes, which typically comprise three digits (letters from RIASEC). In some cases (depending on the specific function), they can also be shorter (one letter) or longer (up to six letters). To make them easier to find, all functions carry the prefix 'con_'

and have a similar structure with regard to their main function arguments. The first two function arguments always refer to the person code (argument a) and to the environment code (argument b).

For a first example, we consider the function `con_oneletter_holland()` and its application in `R_snippet_001` which returns the one letter congruence index according to Holland [28].

R_snippet_001.R

```
1 library(holland) # loads the package ... assuming it is installed
2 con_oneletter_holland(a="RIASEC",b="AIRCES")
3 con_oneletter_holland(a="RIASEC",b="AIRCES",letter = 2)
4 con_oneletter_holland(a="RIASEC",b="RIASEC")
5 con_oneletter_holland(a="RIASEC",b="RIASEC",hexadist = TRUE,letter = 1)
6 con_oneletter_holland(a="RIASEC",b="IRASEC",hexadist = TRUE,letter = 1)
7 con_oneletter_holland(a="RIASEC",b="AIRCES",hexadist = TRUE,letter = 1)
8 con_oneletter_holland(a="RIASEC",b="SIRCEA",hexadist = TRUE,letter = 1)
```

In the simplest default setting (see code line 2), the function allows the computation of a dichotomous congruence index based on the first letter of two Holland letter codes. With the additional argument 'letter', one can control which letter in the two sequences, with a maximum length of 6 letters, the inference of congruence should be based on (see code line 3). Note that unlike in the other 'classical' congruence functions, a larger numerical value (in this case 1 vs. 0) stands for miss fit in congruence here. If the argument 'hexadist=TRUE' is set, this dichotomous congruence statement is extended by a gradual congruence statement whose expression depends on the distance of the respective interest orientation in the hexagonal configuration (see code line 5–8 in `R_snippet_001`). Algorithmically, this is realized internally by establishing a 6×6 distance matrix for the six RIASEC dimensions, where the values 0 stand in the diagonal and the integer distance values (1, 2, 3) stand with increasing distance from the diagonal, where a 3 stands for a letter combination of dimensions opposite in the hexagon. These distances then flow into the calculation of the Holland one-letter congruence index.

In a similar way, the application of the functions `con_twoletter_holland()` and `con_threeletter_holland()` works but without the option to consider the hexagonal arrangement (hexadist) of the dimensions of vocational interests. The respective algorithmic basis of the two indices is given in ([two letter code] Healy and Mourtou [46]) and ([three letter code] Wolfe and Betz [56]).

In the R package *holland*, there are a total of five additional 'classical' congruence indices, as well as two new sequence based indices, whose application is illustrated in `R_snippet_002`.

R_snippet_002.R

```
1 # assuming 'holland' is loaded
2 # create a vector of some possible combinations of Holland three-letter codes -
3 person <- apply((combn(x = c("R","I","A","S","E","C"), m=3)),2,paste0,
4 collapse="")
5 # create a single environment Holland three-letter code -----
6 environ <- "RIA"
7 # create a vector of function names to be called for congruence indices -----
8 func <- c("con_hamming_holland","con_levenshtein_holland","con_brown_c_holland"
9 ,"con_compindex_holland","con_iachan_holland","con_n3_holland",
10 "con_zs_holland")
11 # apply all functions to all combinations of Holland three-letter codes -----
12 con_indices <- sapply(func, function(x){
13   sapply(person, function(y){do.call(what = x, args= list(a=y, b=environ))})
14 })
15 # look at the resulting indices for congruence with "RIA" environment code ----
16 con_indices
```

Code line number 3–4 creates a (character) vector of some possible combinations of Holland three-letter codes, which are taken as the person codes in the subsequent code lines. In code line number 6, a single environment Holland three-letter code is assigned, and in code line number 8–10, a (character) vector of function names for the five 'classical' as well as the two sequence based congruence indices is assigned. In code line number 12–14 of `R_snippet_002`, we apply each of these functions to the vector of the person

codes and the single environment code to return a matrix with two plus five columns, one for each congruence index with values for congruence in each row for the person codes respectively (see Table 2).

Table 2. Values for congruence between the single “RIA” environment code and some possible combinations of personal three-letter codes for seven congruence-indices.

	Sequence Based Indices			‘Classical’ Indices			
	Hamming ^g	Levenshtein ^f	Brown–C ^a	Comp. ^b	Iachan ^c	N3 ^d	ZS ^e
RIA	0.00	0.00	18	8	28	3	6
RIS	0.02	0.02	17	7	27	2	5
RIE	0.50	0.37	16	7	27	2	5
RIC	0.75	0.50	15	7	27	2	5
RAS	0.33	0.13	15	5	24	2	2
RAE	0.81	0.37	14	5	24	2	2
RAC	1.06	0.50	13	5	24	2	2
RSE	1.12	0.40	12	0	22	1	3
RSC	1.37	0.52	11	0	22	1	3
REC	1.68	1.00	9	0	22	1	3
IAS	0.71	0.38	12	4	12	2	2
IAE	1.18	0.62	11	4	12	2	2
IAC	1.43	0.75	10	4	12	2	2
ISE	1.50	0.65	9	3	10	1	1
ISC	1.75	0.77	8	3	10	1	1
IEC	2.06	1.12	6	3	10	1	1
ASE	1.87	1.13	6	2	4	1	1
ASC	2.12	1.26	5	2	4	1	1
AEC	2.43	1.50	3	2	4	1	1
SEC	2.81	2.01	0	0	0	0	0

Notes. ^a Brown–C: Index of Congruence according to Brown and Gore [43]; ^b Comp.: Comparative-Index of Congruence according to Wiggins and Moody [57]; ^c Iachan: Index of Congruence according to Iachan [50]; ^d N3: Index of Congruence according to Joerin Fux [51]; ^e ZS: Index of Congruence according to Zener and Schnuelle [55]; ^f Levenshtein: Sequence based Index according to Levenshtein [62]; ^g Hamming: Sequence based Index according to Hamming [61].

The theoretical foundation and their algorithmic rationale of the five ‘classical’ congruence indices can be found in Brown and Gore [43], Iachan [50], Joerin Fux [51], Zener and Schnuelle [55], Wiggins and Moody [70] respectively.

If we take a closer look at the Table 2 and the resulting values for the ‘classical’ congruence indices (rightmost five columns) it contains (with reference to this example), we notice that the various indices do not come to a mutually consistent (rank) order with regard to the extent of congruence in all cases. Apart from the obvious fact that all indices agree in their assessment of maximum and minimum congruence, there are sometimes considerable differences in the resulting congruence statements. For example, if we compare the resulting rank order of congruence with respect to the three-letter codes for the Iachan-Index and the ZS-Index, we see that the Iachan-Index (compared to the ZS-Index) apparently gives a stronger weighting to the first letter in the three-letter codes. This leads to the fact that the congruence between the letter codes “RAS”, “RAE” and “RAC” and the (environmental) code “RIA” is rated as comparatively high with a value of 24 on a scale of 0 to 28, whereas the codes “IAS”, “IAE” and “IAC” are rated as low with a value of 12 with regard to congruence. In comparison, the ZS-Index does not differentiate at all between these six different letter codes in terms of their degree of congruence, assigning all letter codes the rather low value of 2 in relation to its range of values. With this regard already Iachan [50] criticized the ZS-Index to “not discriminate

sufficiently among the possible outcomes" ([50], p. 136). As can be seen in Table 2, among the 'classical' indices, the most subtle and possibly important differentiation between person-environment congruence is achieved by applying the Brown-C-Index. This index assigns, with few exceptions compared to the other Indices, different congruence values to each of the three letter codes. In detail, in the example demonstrated here, the Brown-C-Index has identical double classifications for congruence with the environmental code "RIA" only for the letter combinations "RIC" and "RAS" (value 15), "RSE" and "IAS" (value 12), "RSE" and "IAE" (value 11), "REC" and "ISE" (value 9), and for the letter codes "IEC" and "ASE" (value 6).

Although the Brown-C-Index differentiates much more finely than the other four classic indices with regard to the degree of congruence, duplicate classifications also occur here. This problem is addressed (to varying degrees) by two novel indices of congruence derived from sequence data analysis (see first two columns in Table 2).

The determination of the congruence of two Holland letter codes can, in addition to the classical indices, also be treated with methods of sequence data analysis. According to Abbott [71], such sequences are defined as ordered lists of elements from a finite selection. This general definition of a sequence or of sequence data also applies to Holland letter codes so that methods from sequence data analysis are an interesting approach for the investigation of the congruence of two Holland letter codes (see [72], [for theoretical overview and empirical investigation]). In the Holland package, two sequence-based distance measures are currently implemented as indices for congruence. One is the Hamming distance [61] and the other is the Levenshtein distance [62].

The function `con_hamming_holland()` computes the location-weighted, cost-sensitive Hamming distance [61]. The function `con_levenshtein_holland()` finds the distance according to Levenshtein [62] between two sequences (see, [71]), which are the Holland codes given in argument `a`, which is the person code, and argument `b`, which is the environment code. Computational details can be found in Needleman and Wunsch [73].

If we look at the results for these two measures of congruence in the first two columns in Table 2, we first notice that they produce an opposite ranking for the person letter codes compared to the other five classical indices. This circumstance can be explained by their different theoretical foundation as distance measures between two code sequences (see [61,71,73]). In addition, it becomes clear that—in contrast to the classical indices—at least with the Hamming distance, in this example considering the fit to the single "RIA" environment code the combinations of the person codes are assigned a unique value for congruence.

Based on the finding from the comparison of the indices presented above we would recommend the following regarding the selection of the suitable congruence index, depending on the respective goal of its application. If for example the main goal is to maximize the precision and uniqueness in differentiation of congruence based on a Holland letter code, applying the Hamming or Levenshtein distance would probably be the best way to go. If on the other hand one would like to stress the influence of the first letter in an interest profile one should consider the use of the Iachan-Index. The focus on the first letter of the interest profile can be based, for example, on the observation of a rather large difference between the highest interest score and the second and third highest.

4.2. Differentiation Indices in holland

With regard to the concept of *differentiation* seven indices as depicted in Table 3 are currently implemented in the package `holland`.

Table 3. Differentiation indices for Holland profiles implemented in the R-package *holland* as cited in ([7], p. 267).

Index	Brief Description	Author/Source
DI1	Difference between highest and second highest interest score	(Frantz & Walsh, 1972) [66]
DI2	Difference between highest and third highest interest score	(Spokane & Walsh, 1978) [67]
DI3	Difference between highest score and the average of the second and fourth highest score	(Iachan, 1984) [68]
DI4	Difference between highest score and the average of the third and fifth highest score	(Iachan, 1984) [68]
DI5	Difference between highest and lowest score	(Holland, 1973) [45]
DI6	Difference between highest and lowest score, standardized by the overall level of interest	(Peiser & Meir, 1978) [69]
DI7	Dispersion of interest scores	(Healy & Mourton, 1983) [46]

These seven indices (see Table 3) are implemented in one single function named 'dif_7_holland' within the R-package *holland*. As to call one of these indices one has to specify which index for differentiation to compute by assigning the respective character expression (as listed in the first column in Table 3) to the argument 'ind' within the function (e.g., ind = "DI5", to compute the differentiation according to [45]). In the subsequent R snippet 'R_snippet_003.R' we first create data for four possible RIASEC score profiles (see code lines 1–5).

R_snippet_003.R

```

1 # create a matrix with possible interest score profiles: -----
2 SP <- matrix(data = c(50,15,15,12,10,10,
3   50,45,15,14,12,40,
4   50,48,26,20,14,10,
5   50,50,50,50,50,50)
6 ,nrow = 4,ncol = 6,byrow = T,dimnames = list(c("Profile_1","Profile_2",
7   "Profile_3","Profile_4"),c("R","I","A","S","E","C")))
8 # plot the four score profiles: -----
9 matplot(x = t(SP),type = "b",pch = " ",xaxt="n",ylim = c(10,50),bty="n",
10  ylab = "raw score",lty=c(1:4))
11 axis(side = 1,at=c(1:6),labels=colnames(SP))
12 segments(x0 = rep(1,5), y0 = seq(20,40,10), x1 = rep(6,5),
13  y1 = seq(20,40,10),col = "gray80")
14 segments(x0 = 1:6, y0 = rep(10,6), x1 = 1:6, y1 = rep(50,6),col = "gray80")
15 text(x = 1:6,y = c(t(SP)), labels = c(t(SP)), cex = .8)
16 legend(1,20,legend = rownames(SP),cex = .7, col = 1:4, lty=c(1:4))
17 # compute differentiation according to Holland (1973) -----
18 apply(SP, 1, dif_7_holland, ind = "DI5")
19 # compute all of the seven differentiation indices -----
20 ind <- c("DI1","DI2","DI3","DI4","DI5","DI6","DI7")
21 sapply(ind, function(x){apply(SP, 1, dif_7_holland, ind = x)})

```

In the code lines 8–16 we give an example on how to plot the four score profiles, which results in a graphical representation of the score profiles as depicted in Figure 3.

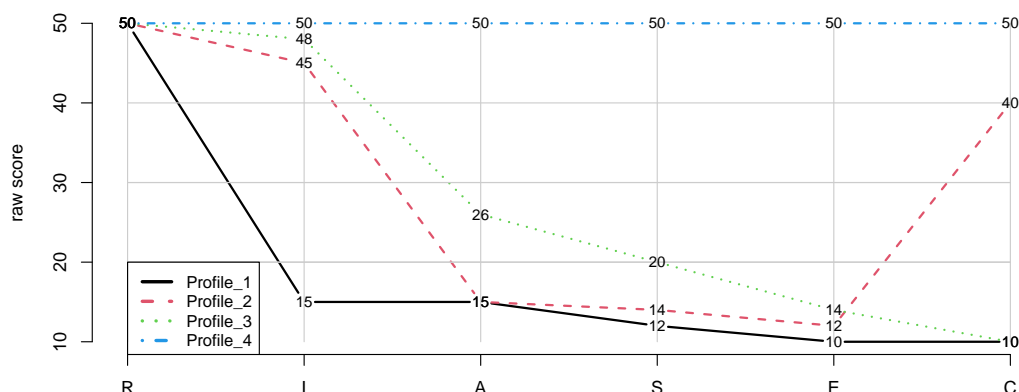


Figure 3. Lineplots for four different RIASEC raw score profiles.

As the Figure 3 illustrates, the four profiles differ on the one hand in terms of their absolute level and on the other hand in terms of their variability of trait expression on the six RIASEC dimensions. For example, Profiles 1 and 3 have the same maximum and minimum (across all six dimensions), but Profile 3 has more variance overall in terms of trait expression across the six dimensions as compared to Profile 1. Note, that all trait values used in this example refer to raw-scores as they result from the application of an RIASEC assessment instrument like the AIST-3 [9] or ASIT-R [8]. In contrast, Profile 4 has the lowest variance with a maximum score on all dimensions. Finally, Profile 2 represents a (typical) clear and consistent 3-letter code profile.

In code line 18 of R snippet 'R_snippet_003.R', we now calculate the differentiation according to Holland [45] for all of these four profiles. As expected, this index ("DI5") cannot distinguish the two profiles 1 and 3 with respect to their differentiation and assigns a value for differentiation of 40 to both (higher values indicate higher differentiation). In order to compare the different ability of the seven indices to discriminate the four score profiles in more detail, we can calculate values for the differentiation according to all seven indices for the four profiles by running code lines 20–21. From the results summarized in Table 4 we see, that the seven indices produce different rankings of the four profiles with regard to their differentiation.

Table 4. Values of seven indices for differentiation on four different types of score profiles.

	DI1	DI2	DI3	DI4	DI5	DI6	DI7
Profile_1	35.00	35.00	36.50	37.50	40.00	0.36	14.16
Profile_2	5.00	10.00	20.00	23.00	38.00	0.22	15.95
Profile_3	2.00	24.00	16.00	30.00	40.00	0.24	15.66
Profile_4	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Notes. The values for differentiation base on a sum score scale ranging from 10 to 50 for each dimension respectively.

For example, the measures "DI1", "DI2", "DI3", "DI4", and "DI6" each assign different scores regarding Profile 1 and Profile 3 and indicate that Profile 1 shows the highest differentiation. In contrast, according to "DI7", Profile 2 is the most differentiated one. However, all indices agree that Profile 4 is the least differentiated interest profile. Since the measures "DI1" to "DI6" do not take into account all RIASEC scores equally, "DI7", the dispersion of the six RIASEC scores, appears to be the most sophisticated measure of Holland's concept of differentiation.

4.3. The Calculus Hypothesis in holland

Given that the user has a version > 8.4 of the commercial program Mplus installed on her or his computer, a functional area of holland can be used, which is related to the calculus hypothesis and allows the analysis of the empirical hexagonal arrangement, the testing for fit of the six dimensions of interest, as well as their graphical representation. The functions

make use of the principle of estimating the angular locations of the six RIASEC dimensions by means of a structural equation modeling approach implemented as *Mplus* Syntax by Nagy et al. [18]. Due to the fact that these are functions which more or less only control an external software whose existence on the computer cannot be assumed by all R users, we will not go into a detailed presentation of all options possible in the related functions here, but only briefly outline the use of the three core functions of this area with two associated plotting S3 methods. In addition we will present a short R script which was used to produce the projection plot presented in the Figure 2 in the section related to the introduction into Holland's theory.

Basically there are three functions, which all create *Mplus* executable syntax code, in turn automatically run *Mplus* on this code and finally read the respective result into R again. The function `Circ_emp()` allows for computing the *empirical* angular locations of the six interest dimensions based on a full correlation matrix and the respective sample size. Some hints on how to structure such a full correlation matrix are given in some examples, which can be accessed by typing the R command `data(example1)` or `data(example2)`. The result of executing this function can be assigned to an related plotting S3 Method to return a graphical visualization of the estimated angular locations. The subsequent R snippet demonstrates how to produce the Figure 2 visualizing some empirical RIASEC correlations initially analyzed by Heine et al. [24].

R_snippet_004.R

```

1 ## assuming 'holland' is loaded and Mplus is installed
2 # (re)construct the correlation matrix from Heine, Langmeyer & Tarnai (2011) --
3 korr <- matrix(data = c(1.0000,0.5411,-0.0882,-0.1294,-0.1329,0.1417,0.5411,
4   1.0000,0.2105,0.0045,-0.0440,0.1047,-0.0882,0.2105,
5   1.0000,0.4593,0.1905,0.0914,-0.1294,0.0045,0.4593,
6   1.0000,0.4692,0.2288,-0.1329,-0.0440,0.1905,0.4692,
7   1.0000,0.4540,0.1417,0.1047,0.0914,0.2288,0.4540,
8   1.0000),nrow = 6,ncol = 6,byrow = F); korr
9 # generate, run and read in results from Mplus code -----
10 result <- Circ_emp(N = 734,Cor = korr,konstrukt = c("R","I","A","S","E","C"))
11 # plot the result object -----
12 plot(result,lcolor = c("black","black","blue","blue"),ltype = c(1, 1),lwd = 3,
13   defhexa = list(hexa = TRUE, seg = TRUE, gr = 5, r = 4, nseg = 7,
14     x.cent = 0, y.cent = 0, circle=T) )
15 # add a grey circle around the hexagon -----
16 r <- 4.05
17 Hxx <- r*sin( seq(0),(2*pi), length.out=360) )
18 Hyy <- r*cos( seq(0),(2*pi), length.out=360) )
19 lines(Hxx,Hyy, col="grey90",lty=1,lwd=3)

```

Using the function `Circ_test()`, it is possible to test whether any given full correlation matrix fits to a given (hexagonal) angular arrangement. The given (hexagonal) angle arrangement to be tested against can be specified with the argument `test` in two different ways. The typical approach is certainly to test whether a given empirical correlation matrix fits the ideal hexagonal structure as postulated in the calculus hypothesis. This is achieved by specifying the argument according to the following expression: `test="perfect"`, which is also the default setting. In addition, it is possible to test against an arbitrary (hexagonal) angle arrangement by passing the six angles (in RIASEC order) as numerical values in radians to the argument `'test'`. The last function `Circ_pro()` allows for the angular projection of additional dimension into the hexagonal structure of the vocational interest dimensions from (possibly) associated constructs, such as for example the Big Five dimensions of personality. For such a projection of additional construct dimensions, an appropriately populated correlation matrix must be passed to the function. How such a correlation matrix has to be structured can be taken from a corresponding data example, which can be accessed via the R command `data(example3)` or `data(example4)`. Finally, the result of the function `Circ_pro()` can be assigned to an related plotting S3 Method to return a graphical visualization of the estimated angular locations for both, the six RIASEC dimensions, as well as the additional construct dimensions, projected into the hexagon.

4.4. The Functional Area Miscellaneous in holland

Under the keyword 'misc', a branch of functions are subsumed, that allow for some convenient data manipulation functionalities. The function `kormean()` takes the mean of the entries of two correlation matrices using the Fisher-Z transformation of the coefficients in both matrices. The function `sco2let()` allows for converting RIASEC score profiles to Holland letter-codes. The function `sim_score_data()` will simulate Person (raw)-scores for an arbitrary number of dimensions (latent variables) assessed with any type of questionnaire given the maximum and minimum raw score for each dimension, and some more functions, just to mention a few important ones. Next to such helper functions there are some internal functions providing a comprehensive plotting environment, which is currently used only by one central plotting function exported to the user. The function `plot_profile_holland()` basically allows for visualizing a individual interest profile given in the form of six score values, which can either be norm values or raw-scores as returned by a vocational interest inventory such as the AIST-R [8] or the AIST 3 [9]. In the subsequent R-snippet ('R_snippet_005'), we first simulate some data using an empirical correlation matrix (code lines 3–7) from the german AIST female norm sample [8], run some descriptives on the resulting data (code lines 9–11), demonstrate two examples to identify ties when ranking RIASEC score profiles (code lines 13–16), apply a conversion to Holland letter-codes (code lines 18–20), and finally add some of the indices for congruence and differentiation (code lines 22–29) already discussed above.

R_snippet_005.R

```

1 ## assuming 'holland' is loaded
2 # get an RIASEC correlation matrix -----
3 data(AIST_2005_F_1270)
4 # simulate raw scores with minimum = 10 and maximum = 50 -----
5 set.seed(1234)
6 D <- sim_score_data(n=1000,cormat=AIST_2005_F_1270,min.score = 10,
7   max.score = 50,data.frame = T)
8 # look at some descriptives -----
9 apply(D, 2, range)
10 apply(D, 2, mean)
11 apply(D, 2, sd)
12 # add index for ties on the 6 dimensions -----
13 D$Ties6 <- apply(D, 1, function(x){length(unique(x[1:6]))!=6})
14 # add index for ties on three highest dimensions -----
15 D$Ties3 <- apply(D, 1, function(x){length(unique(x[1:6][order(x[1:6],
16   decreasing = T)[1:3]))!=3})
17 # add a (character) vector of full 6-letter-codes for every person -----
18 D$Letter6 <- apply(D, 1, function(x){sco2let(x[1:6],len = 6)})
19 # add a (character) vector of 3-letter-codes for every person -----
20 D$Letter3 <- apply(D, 1, function(x){sco2let(x[1:6])})
21 # add the 'DI7' index to data set 'D' -----
22 D$DI7 <- apply(D[,1:6], 1, function(x){dif_7_holland(A = x,ind = "DI7")})
23 # add the 'Hamming distance' index to data set 'D' -----
24 env <- "RIA"
25 D$hamming <- sapply(D$Letter3, function(x){con_hamming_holland(a = x,b = env)})
26 # add the 'Iachan' index to data set 'D' -----
27 D$iachan <- sapply(D$Letter3, function(x){con_iachan_holland(a = x,b = env)})
28 # add the 'Brown-C' index to data set 'D' -----
29 D$brownc <- sapply(D$Letter3, function(x){con_brown_c_holland(a = x,b = env)})
30 head(D,n = 7)

```

Executing the last command line in 'R_snippet_005' results in an output showing the first seven cases in the simulated data set on the R console, as depicted in the box below. Next to the raw scores for each of the six RIASEC dimensions (columns 1–6), there are two index variables pointing to cases with ties on the individual RIASEC rank order (columns 7–8), as well as letter-codes with either a length of six or three letters (columns 9–10), the "DI7" index for differentiation and in the last three columns indices for congruence according to an environment characterized by the code "RIA".

R	I	A	S	E	C	Ties6	Ties3	Letter6	Letter3	DI7	hamming	iachan	brownc	
1	22	30	36	15	33	32	FALSE	FALSE	AECIRS	AEC	7.234178	2.4375	4	3
2	22	22	24	21	25	21	TRUE	FALSE	EARISC	EAR	1.500000	1.5625	6	8
3	23	28	35	28	28	23	TRUE	TRUE	AISERC	AIS	4.031129	0.7750	9	11
4	28	44	33	27	31	33	TRUE	TRUE	IACERS	IAC	5.557777	1.4375	12	10
5	20	18	30	21	28	22	FALSE	FALSE	AECSRI	AEC	4.336537	2.4375	4	3
6	31	25	23	22	19	21	FALSE	FALSE	RIASCE	RIA	3.818813	0.0000	28	18
7	12	15	27	26	36	21	FALSE	FALSE	EASCIR	EAS	7.987838	1.0875	2	9

As can be seen in the output box above, there are some cases which have ties when trying to rank order their individual interest scores on the six RIASEC dimensions. For example, case number 2 shows a tie when ranking is based on all of the six dimensions, but shows no tie when restricting the ranking on the three dimensions with the most highest raw scores only. Therefore, if any further analyses and interpretations are based only on the three-letter code, meaningful and clear analysis statements could be made even for cases like case number 2. However, as with case number 4, ties also occur within the first three letters or between the third and fourth letter. This is especially true with respect to short versions of interest inventories as they are used in large-scale studies, since a reduced number of items lead to a reduced number of different possible scale scores. Because there is no satisfying solution for the problem of ties, methods using full profile information instead of high-point codes to describe a person or an environment in the framework of Holland's theory are recommended [30]. However, there are countries such as Germany where no RIASEC score profiles but only three-letter codes are available to describe work environments (It is possible to use the RIASEC profiles of the O*NET to describe work environments using RIASEC scores, but this should be done with care, since they were created with respect to the US labor market). One method to deal with ties could be to delete all cases with tied RIASEC scores, which could be associated with a substantial reduction of the sample size. Alternatively, tied scores can be ranked randomly or for each case with ties all possible three-letter codes can be used for further analyses (as a consequence, the respective results must be aggregated, e.g., the mean congruence of multiple three-letter codes with regard to one work environment has to be calculated). Similarly, in the context of career counseling, clients with tied RIASEC scores within the first three (or four) letters are advised to consider occupations for all possible three-letter code combinations. For example, case number 4 would be advised to consider occupations that fit to the three-letter codes IAC and ICA.

As mentioned above, differentiation and congruence are supposed to predict favorable outcomes such as satisfaction or performance. To test these hypotheses correlations can be calculated or regression analyses can be applied with DI7 and the hamming distance as independent variables, for example (see also [25,74,75]).

In the last R snippet ('R_snippet_006') we first calculate some descriptives for three congruence indices for the simulated data by sub setting the data using the indicator for ties considering the three strongest dimensions (code lines 4–16). We then demonstrate how to plot a circular representation of a score profile. Here we do not aim at plotting individual score profiles, but (mean) aggregated score profiles for the sub sample of cases showing congruence above vs. below the mean of the congruence index distribution respectively (code lines 20–30). These examples show how different samples can be described based on the interests of the respondents. In addition, in this context, the differences in the congruence indices used can be clearly illustrated.

R_snippet_006.R

```

1 ## assuming 'holland' is loaded and data 'D' is present in workspace
2 head(D,n = 7)
3 # calculate means for congruence indices by ties vs no ties -----
4 by(data=D,INDICES=D$Ties3,function(x){
5   colMeans(x[,c("hamming","iachan","brownc")])
6 })
7 # calculate standard deviation for congruence indices by ties vs no ties -----
8 by(data=D,INDICES=D$Ties3,function(x){
9   sapply(c("hamming","iachan","brownc"),function(y){
10    sd(x[,y])
11  })
12 })
13 # calculate correlations between congruence indices by ties vs no ties -----
14 by(data = D,INDICES = D$Ties3, function(x){
15   cor(x[,c("hamming","iachan","brownc")])
16 })
17 # sub setting data restricted to cases without ties on 3 dimensions -----
18 D1 <- D[D$Ties3==FALSE,]
19 # calculate score profile means below vs. above mean iachan congruence -----
20 iac_UPm <- colMeans(D1[D1$iachan>=mean(D1$iachan),1:6])
21 iac_LOm <- colMeans(D1[D1$iachan<=mean(D1$iachan),1:6])
22 # plot circular score profile means below vs. above mean iachan congruence ----
23 plot_profile_holland(x=iac_UPm,ri.M=10,ro.M=50,cex.sl=.8,cex.la=1.4,circle=T)
24 plot_profile_holland(x=iac_LOm,ri.M=10,ro.M=50,cex.sl=.8,cex.la=1.4,circle=T)
25 # calculate score profile means below vs. above mean hamming congruence -----
26 ham_UPm <- colMeans(D1[D1$hamming>=mean(D1$hamming),1:6])
27 ham_LOm <- colMeans(D1[D1$hamming<=mean(D1$hamming),1:6])
28 # plot circular score profile means below vs. above mean hamming congruence ---
29 plot_profile_holland(x=ham_UPm,ri.M=10,ro.M=50,cex.sl=.8,cex.la=1.4,circle=T)
30 plot_profile_holland(x=ham_LOm,ri.M=10,ro.M=50,cex.sl=.8,cex.la=1.4,circle=T)

```

Note, that for the sake of distinctiveness, the plots of the (mean) aggregated score profiles are based on the data of those cases only for which there are no ties of rank in the RI-ASEC profile based on the three strongest dimensions (see code line 18 in 'R_snippet_006').

From the four panels in Figure 4 we see that although the mean aggregated score profiles have a almost similar shaped polygon area, the resulting overall vector for the two sub-samples (below and above the mean for congruence, respectively) point in significantly different directions. For the group of above average congruent cases to the environment "RIA", the mean resulting vector lies in the first quadrant between "R" and "I". The comparison of the two indices (panel **b** and **d**) for this most congruent sub sample shows that for the Iachan index (panel **b**) with $V = 19.22^\circ$ a smaller angle is shown than for the Hamming distance (panel **d**) with $V = 33.39^\circ$. This finding reflects the fact that, as described above, the Iachan index emphasizes the first letter of the 3-letter code more strongly, and thus the mean overall vector approaches the dimension *Realistic*. Correspondingly, regarding the sub samples of below average congruent cases, the vectors shows in the direction of the Social dimension that is antagonistic to the Realistic dimension. Again, this is more pronounced when the Iachan index is used ($V = 171.38^\circ$) instead of the Hamming distance ($V = 160.96^\circ$).

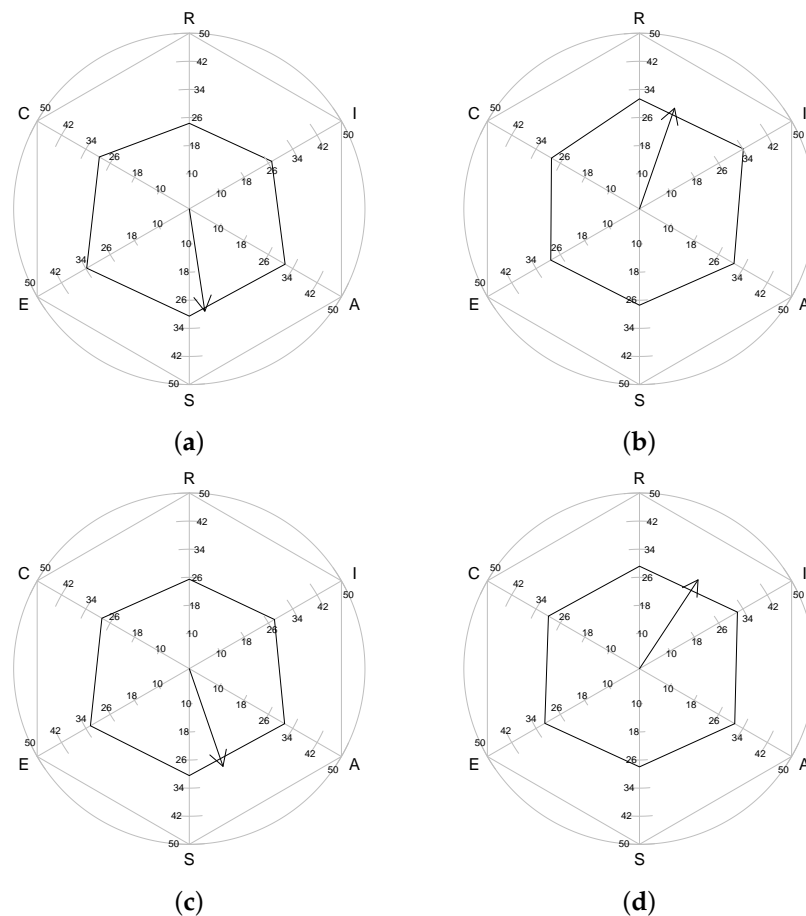


Figure 4. Circular graphical visualizations for mean aggregated interests score profiles with resulting mean interests vector V ; upper panel (a,b): Subsamples mean splitted by Iachan [50] index, ((a): below mean, $V = 171.38^\circ$, (b): above mean, $V = 19.22^\circ$); lower panel (c,d): Subsamples mean splitted by Hamming distance [61] index, ((c): above mean, $V = 160.96^\circ$, (d): below mean, $V = 33.39^\circ$); all values in degree clockwise starting at dimension *Realistic*.

5. Discussion and Conclusions

The R package *holland* has recently been released by our research group on the repositories of the comprehensive R Network (CRAN) and provides an initial framework for implementing the theory of vocational choice. It aims at providing a comprehensive framework to determine the coefficients that quantify the central concepts of Holland's [1] theory in a convenient way. Although Holland's theory is more than 60 years old, it is still one of the most cited and most applied vocational approaches all over the world [4,6]. Getting started with the theory and its application can be difficult, especially because there are an overwhelming number of ways to quantify the same concepts, which are described in numerous different works. Therefore, the R package *holland* can become an important tool that comprises all relevant coefficients based on the RIASEC model and, consequently, eases the everyday work of researchers and practitioners in vocational psychology. In its current state, it has implemented many of the central concepts and coefficients that can be found in the vocational literature. However, some concepts and some coefficients are not provided yet. There might also be potential for improvement in terms of user-friendliness for practitioners such as career counselors.

Regarding the concept of congruence, the R package *holland* focuses on measures that are based on high-point codes. However, these measures have some flaws. First of all, they only consider the stronger RIASEC dimensions and ignore relevant information given by the weaker RIASEC dimensions [29]. Second, they do not take the differences between the dimensions of a high-point code into account. For example, if *Realistic* is the

dominant dimension followed by Investigative, a small difference between the scores of these two dimensions (e.g., 2 points) is treated equally to a bigger difference (e.g., 20 points). However, such differences may have an effect on vocational behavior and the fit to a work environment. Third, because ties are possible the RIASEC dimensions can not always be clearly ranked and there is no unequivocal method to deal with this problem [25,42]. In addition, profile-based measures of congruence seem to have a better criterion validity than the congruence indices that use high-point codes [42]. Although congruence measures based on high-point codes have weaknesses, it may not always be possible to use full RIASEC score profiles; for example, there may be a case where only environmental high-point codes are available. For these cases, the R package *holland* provides well known and established measures, as well as some interesting new measures, that have rarely been considered in the vocational literature such as the Hamming distance [61] or the Levenstein distance [62]. These measures seem to be promising in terms of depicting the hexagonal structure of vocational interests [72]; however, proof of the criterion validity still has to be provided.

To determine congruence, personal and environmental RIASEC profiles are required. This means that RIASEC profiles, ideally full RIASEC score profiles, have to be assigned to the occupations the respondents have indicated. A manual assignment of occupational RIASEC codes needs a lot of resources and is error-prone, especially when large-scale datasets are used. Here, the information about the respondents' occupations is usually provided as occupational codes belonging to an (inter)national occupational classification, such as the International Standard Classification of Occupations (ISCO) codes or the German Klassifikation der Berufe (KldB) codes to give an example for a national classification. A time-saving way to assign occupational RIASEC profiles is to use such occupational codes and link them to RIASEC profiles that are provided by databases such as the O*NET. Therefore, the future development of the R package should provide a crosswalk from the O*NET SOC codes and their occupational RIASEC profiles to the ISCO codes to which national occupational codes can easily be linked. Although the occupational RIASEC profiles of the O*NET have been developed for the United States, this may be the only opportunity for other nations to efficiently assign RIASEC score profiles to occupations in a large-scale dataset.

The concept of consistency is not yet covered by the R package *holland*. Here, the measure by Tracey should especially be implemented since it is based on full RIASEC score profiles and, in contrast to high-point code-based measures, predicts important career-related outcomes [25].

The concept of differentiation is well implemented in the R package providing many different coefficients that are based on scores. Future versions of the package could also use Tracey's cosine based approach that takes full RIASEC score profile information and the hexagonal assumption into account [25].

As with consistency, the concept of interest flexibility (profile elevation) is not yet covered by the R package. But this can be implemented directly in R with very simple ways, and moreover, it might be easy to implement in an upcoming version.

In addition to the operationalization of Holland's concept, the R package *holland* provides a sophisticated method to evaluate the calculus hypothesis. An interesting feature here is that other variables such as the Big Five can be projected into the perfect hexagon together with the RIASEC dimensions illustrating their relation to each other [19]. A weakness regarding the evaluation of the calculus hypothesis in the current version of the R package is that a paid program (Mplus) is additionally required. Alternatively, the RTOHOR [16] as a manifest-variable approach could easily be integrated since Tracey has already written an R code for this test (see <https://isearch.asu.edu/profile/229363>, accessed on 20 November 2021).

Within the functional area miscellaneous, the plot function seems to be especially interesting for counselors and their clients as it enables a quick illustration of the vocational

interests of a person (or environment). An additional feature could be to superimpose a personal and an environmental profile so that similarities and differences become visible.

The fact that the package offers many different coefficients may be helpful for scientists, who, for example, aim at carrying out comparative studies. However, the methodological diversity may also be confusing for practitioners and their clients since different measures may produce different career recommendations for the same personal interest profile [3,42]. Therefore, future developments of the R package *holland* could implement a collection of coefficients that provide only those coefficients that are preferred according to the current state of science. This collection should be retrievable with one command only and lead to a clear result in terms of congruence, consistency, differentiation, and interest flexibility and should also contain an illustration that depicts the client's RIASEC profile (see Figure 4). If the O*NET (or a national list of occupations with assigned RIASEC profiles) is also integrated here, occupational RIASEC profiles can additionally be illustrated for comparison and recommendations regarding promising career decisions can directly be given.

To conclude, the R package *holland*, in its current state, covers the most important concept of Holland's [1] theory, namely congruence. The concept of congruence is widely studied [21,23] and can also be considered as the main goal of career counseling using a person-environment fit approach. In addition, the package provides measures that quantify the concept of differentiation, which relates to important outcomes [25] and is therefore of relevance for both researchers and practitioners. However, future versions of the package should also include the implementation of measures that quantify the concepts of consistency and elevation. Regarding congruence, some measures that are based on full RIASEC score profiles should be considered [42].

Holland's theory has undergone some modifications and the addition of several coefficients and indices in its now almost 70 year history. It is therefore hardly unusual that the only R-package now available for Holland's theory does not instantly represent all concepts and coefficients in the form of corresponding R-functions. Rather, the R package now provides a sound and modularly extensible framework into which future extensions can be easily inserted. Not least the feedback from users, for whom we have given an introduction to the theory as well as to the practical application of the R-package with this contribution, will be able to significantly influence the direction of future extensions.

Author Contributions: Conceptualization, J.-H.H., F.G.H. and B.E.; methodology, F.G.H., J.-H.H. and B.E.; software, J.-H.H. and F.G.H.; validation, F.G.H., J.-H.H., and B.E.; writing—original draft preparation, F.G.H. and J.-H.H.; writing—review and editing, B.E., F.G.H. and J.-H.H.; visualization, J.-H.H.; supervision, J.-H.H. and F.G.H.; project administration, F.G.H. and J.-H.H.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: We would like to thank Diana Lee Sosa for the language revision of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Holland, J.L. *Making Vocational Choices. A Theory of Vocational Personalities and Work Environments*, 3rd ed.; Psychological Assessment Resources: Odessa, FL, USA, 1997.
2. Juntunen, C.L.; Motl, T.C.; Rozzi, M. Major career theories: International and developmental perspectives. In *International Handbook of Career Guidance*; Athanassou, J.A., Perera, H.N., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 45–72. [CrossRef]
3. Su, R.; Murdock, C.; Rounds, J. Person-environment fit. In *APA Handbook of Career Intervention, Volume 1: Foundations*; Hartung, P.J., Savickas, M.L., Walsh, W.B., Eds.; APA Handbooks in Psychology®; American Psychological Association: Washington, DC, USA, 2015; pp. 81–98. [CrossRef]

4. Kennelly, E.; Sargent, A.; Reardon, R. *RIASEC Literature from 1953-2016: Bibliographic References to Holland's Theory, Research, and Applications (Technical Report No. 58)*; Technical Report; Florida State University: Tallahassee, FL, USA 2018; doi:10.17125/fsu.1519670557. [CrossRef]
5. Holland, J.L.; Fritzsche, B.A.; Powell, A.B. *The Self-Directed Search Technical Manual*; Psychological Assessment Resources: Odessa, FL, USA, 1994.
6. Hansen, J.I.C. Remembering John L. Holland, PhD. *Couns. Psychol.* **2011**, *39*, 1212–1217. [CrossRef]
7. Bergmann, C. Differenziertheit der Interessen und berufliche Entwicklung. *Z. Differ. Diagn. Psychol.* **1993**, *14*, 265–279.
8. Bergmann, C.; Eder, F. *AIST-R Allgemeiner Interessen-Struktur-Test mit Umwelt-Struktur-Test (UST-R)-Revision*; Beltz: Göttingen, Germany, 2005.
9. Bergmann, C.; Eder, F. *AIST 3 Allgemeiner Interessen-Struktur-Test mit Umwelt-Struktur-Test (UST-3)-Version 3*; Hogrefe: Göttingen, Germany, 2019.
10. Joerin, S.; Stoll, F.; Bergmann, C.; Eder, F. *EXPLORIX. Das Werkzeug zur Berufswahl und Laufbahnplanung*; Hans Huber: Göttingen, Germany, 2004.
11. Heine, J.H.; Hartmann, F.G. *Holland: Statistics for Holland's Theory of Vocational Choice*; R Package Version 0.1.2-1. 2021. Available online: <https://cran.r-project.org/web/packages/holland/index.html> (accessed on 20 November 2021)
12. Eder, F.; Bergmann, C. Das Person-Umwelt-Modell von J. L. Holland. Grundlagen-Konzepte-Anwendungen. In *Berufliche Interessen. Beiträge zur Theorie von J. L. Holland*; Tarnai, C., Hartmann, F.G., Eds.; Waxmann: Münster, Germany; New York, NY, USA, 2015; pp. 11–30.
13. Hartmann, F.G. *Analysen zur Übereinstimmung Beruflicher Interessen in der Familie. Eine Studie zur Bestimmung der Dyadischen Ähnlichkeit bei Mehrdimensionalen Konstrukten im Familialen Kontext*; Empirische Erziehungswissenschaft, Waxmann: Münster, Germany, 2018.
14. Rounds, J.; Tracey, T.J.; Hubert, L. Methods for evaluating vocational interest structural hypotheses. *J. Vocat. Behav.* **1992**, *40*, 239–259. [CrossRef]
15. Tracey, T.J.; Rounds, J.B. Evaluating Holland's and Gati's vocational-interest models: A structural meta-analysis. *Psychol. Bull.* **1993**, *113*, 229–246. [CrossRef]
16. Hubert, L.; Arabie, P. Evaluating order hypotheses within proximity matrices. *Psychol. Bull.* **1987**, *102*, 172–178. [CrossRef]
17. Tracey, T.J.G. RANDALL: A microsoft FORTRAN program for a randomization test of hypothesized order relations. *Educ. Psychol. Meas.* **1997**, *57*, 164–168. [CrossRef]
18. Nagy, G.; Marsh, H.W.; Lüdtke, O.; Trautwein, U. Representing the circles in our minds: Confirmatory factor analysis of circumplex structures and profiles. In *Structural Equation Modeling in Educational Research-Concepts and Applications*; Teo, T., Khine, M.S., Eds.; Contemporary Approaches to Research in Learning Innovations, Sense Publishers: Rotterdam, The Netherlands; Boston, MA, USA; Taipei, Taiwan, 2009; pp. 287–316.
19. Hurtado Rúa, S.M.; Stead, G.B.; Poklar, A.E. Five-Factor personality traits and RIASEC interest types: A multivariate meta-analysis. *J. Career Assess.* **2019**, *27*, 527–543. [CrossRef]
20. Donohue, R. Person-environment congruence in relation to career change and career persistence. *J. Vocat. Behav.* **2006**, *68*, 504–515. [CrossRef]
21. Hoff, K.A.; Song, Q.C.; Wee, C.J.M.; Phan, W.M.J.; Rounds, J. Interest fit and job satisfaction: A systematic review and meta-analysis. *J. Vocat. Behav.* **2020**, *123*, 103503. [CrossRef]
22. Nye, C.D.; Su, R.; Rounds, J.; Drasgow, F. Vocational interests and performance. A quantitative summary of over 60 years of research. *Perspect. Psychol. Sci.* **2012**, *7*, 384–403. [CrossRef]
23. Nye, C.D.; Su, R.; Rounds, J.; Drasgow, F. Interest congruence and performance: Revisiting recent meta-analytic findings. *J. Vocat. Behav.* **2017**, *98*, 138–151. [CrossRef]
24. Heine, J.H.; Langmeyer, A.; Tarnai, C. Auswirkung individueller Antwortstile auf die hexagonale Modellstruktur der Beruf-sinteressen. In Proceedings of the 76th Conference of the Arbeitsgruppe für Empirische Pädagogische Forschung (AEPF); Presentation at the 76th Conference of the Arbeitsgruppe für Empirische Pädagogische Forschung (AEPF), Klagenfurt, Austria, 5–7 September 2011.
25. Tracey, T.J.G.; Wille, B.; Durr, M.R., II; De Fruyt, F. An enhanced examination of Holland's consistency and differentiation hypotheses. *J. Vocat. Behav.* **2014**, *84*, 237–247. [CrossRef]
26. Gupta, A.; Chong, S.; Leong, F.T.L. Development and validation of the vocational identity measure. *J. Career Assess.* **2015**, *23*, 79–90. [CrossRef]
27. Darcy, M.; Tracey, T.J.G. Integrating abilities and interests in career choice: Maximal versus typical assessment. *J. Career Assess.* **2003**, *11*, 219–237. [CrossRef]
28. Holland, J.L. Explorations of a theory of vocational choice and achievement: II. A four-year prediction study. *Psychol. Rep.* **1963**, *12*, 547–594. [CrossRef]
29. Dik, B.J.; Strife, S.R.; Hansen, J.I.C. The flip side of Holland type congruence. Incongruence and job satisfaction. *Career Dev. Q.* **2010**, *58*, 352–358. [CrossRef]
30. Tracey, T.J.G.; Sodano, S.M. Structure of interests and competence perceptions. In *Handbook of Vocational Psychology. Theory, Research, and Practice*; Walsh, W.B., Savickas, M.L., Hartung, P.J., Eds.; Talyor & Francis: New York, NY, USA, 2013; pp. 104–120.

31. Joerin Fux, S.; Stoll, F.; Bergmann, C.; Eder, F. *EXPLORIX.de. Das Werkzeug zur Berufswahl und Laufbahnplanung. Deutschsprachige Adaption und Weiterentwicklung des Self-Directed Search (SDS) nach John L. Holland. 4., Vollständig überarbeitete Auflage. Test-Set Ausgabe Deutschland*; Huber: Bern, Switzerland, 2012.
32. Tracey, T.J.G.; Ward, C.C. The structure of children's interests and competence perceptions. *J. Couns. Psychol.* **1998**, *45*, 290–303. [[CrossRef](#)]
33. Rounds, J.; Smith, T.; Hubert, L.; Lewis, P.; Rivkin, D. *Development of Occupational Interest Profiles for O*NET*; Technical Report; National Center for O*NET Development: Raleigh, NC, USA, 1999.
34. Wille, B.; Tracey, T.J.G.; Feys, M.; De Fruyt, F. A longitudinal and multi-method examination of interest-occupation congruence within and across time. *J. Vocat. Behav.* **2014**, *84*, 59–73. [[CrossRef](#)]
35. O*NET. O*NET OnLine. 2021. Available online: <https://www.onetonline.org/> (accessed on 12 August 2021).
36. Rounds, J.; Su, R.; Lewis, P.; Rivkin, D. *Occupational Interest Profiles for New and Emerging Occupations in the O*NET System: Summary*; National Center for O*NET Development: Raleigh, NC, USA, 2013.
37. Ertl, B.; Hartmann, F.G. The interest profiles and interest congruence of male and female students in STEM and non-STEM fields. *Front. Psychol.* **2019**, *10*, 897. [[CrossRef](#)]
38. Ghetta, A.; Hirschi, A.; Herrmann, A.; Rossier, J. A psychological description of the Swiss labor market from 1991 to 2014: Occupational interest types, sex, salary, and skill level. *Swiss J. Psychol.* **2018**, *77*, 83–94. [[CrossRef](#)]
39. Executive Office of the President, Office of Management and Budget. *Standard Occupational Classification Manual*; Bureau of Labor Statistics: Washington, DC, USA, 2018.
40. Eder, F. Differenziertheit der Interessen als Prädiktor der Interessenentwicklung. In *Pädagogisch-Psychologische Interessenforschung in Studium und Beruf*; Abel, J., Tarnai, C., Eds.; Waxmann: Münster, Germany, 1998; pp. 63–77.
41. Prediger, D.J. Dimensions underlying Holland's hexagon: Missing link between interests and occupations? *J. Vocat. Behav.* **1982**, *21*, 259–287. [[CrossRef](#)]
42. Xu, H.; Li, H. Operationalize interest congruence: A comparative examination of four approaches. *J. Career Assess.* **2020**, *28*, 571–588. [[CrossRef](#)]
43. Brown, S.D.; Gore, P.A. An Evaluation of Interest Congruence Indices: Distribution Characteristics and Measurement Properties. *J. Vocat. Behav.* **1994**, *45*, 310–327. [[CrossRef](#)]
44. Camp, C.C.; Chartrand, J.M. A comparison and evaluation of interest congruence indices. *J. Vocat. Behav.* **1992**, *41*, 162–182. [[CrossRef](#)]
45. Holland, J.L. *Making Vocational Choices: A Theory of Careers*; Prentice-Hall: Englewood Cliffs, NJ, USA, 1973.
46. Healy, C.C.; Mourton, D.L. Derivatives of the Self-Directed Search: Potential clinical and evaluative uses. *J. Vocat. Behav.* **1983**, *23*, 318–328. [[CrossRef](#)]
47. Tarnai, C.; Hartmann, F.G.; Heine, J.H. Ipsative und normative Messung beruflicher Interessen nach dem Holland-Model. In *Klassifikationsanalysen in den Sozialwissenschaften*; Reinecke, J., Tarnai, C., Eds.; Waxmann: Münster, Germany, 2021; pp. 25–52.
48. Eggerth, D.E.; Andrew, M.E. Modifying the C index for use with Holland codes of unequal length. *J. Career Assess.* **2006**, *14*, 267–275. [[CrossRef](#)]
49. Gore Jr, P.A.; Brown, S.D. Simpler may still be better: A reply to Eggerth and Andrew. *J. Career Assess.* **2006**, *14*, 276–282. [[CrossRef](#)]
50. Iachan, R. A measure of agreement for use with the holland classification system. *J. Vocat. Behav.* **1984**, *24*, 133–141. [[CrossRef](#)]
51. Joerin Fux, S. *Persönlichkeit und Berufstätigkeit: Theorie und Instrumente von John Holland im Deutschsprachigen Raum unter Adaptation und Weiterentwicklung von Self-Directed Search (SDS) und Position Classification Inventory (PCI)*; Cuvillier: Göttingen, Germany, 2005.
52. Nye, C.D.; Prasad, J.; Bradburn, J.; Elizondo, F. Improving the operationalization of interest congruence using polynomial regression. *J. Vocat. Behav.* **2018**, *104*, 154–169. [[CrossRef](#)]
53. Tracey, T.J.G.; Robbins, S.B. The interest-major congruence and college success relation: A longitudinal study. *J. Vocat. Behav.* **2006**, *69*, 64–89. [[CrossRef](#)]
54. Fisher, G.A.; Heise, D.R.; Bohrnstedt, G.W.; Lucke, J.F. Evidence for extending the circumplex model of personality trait language to self-reported moods. *J. Personal. Soc. Psychol.* **1985**, *49*, 233–242.
55. Zener, T.B.; Schnuelle, L. Effects of the self-directed search on high school students. *J. Couns. Psychol.* **1976**, *23*, 353–359. [[CrossRef](#)]
56. Wolfe, L.K.; Betz, N.E. Traditionality of choice and sex-role identification as moderators of the congruence of occupational choice in college women. *J. Vocat. Behav.* **1981**, *18*, 43–55. [[CrossRef](#)]
57. Wiggins, J.D.; Moody, A. *Compatibility Index Description*; Training and Associates: Dover, DE, USA, 1981.
58. Robbins, P.I.; Thomas, L.; Harvey, D.W.; Kandefer, C. Career change and congruence of personality type: An examination of DOT-derived work environment designations. *J. Vocat. Behav.* **1978**, *13*, 15–25. [[CrossRef](#)]
59. Kwak, J.C.; Pulvino, C.J. A mathematical model for comparing Holland's personality and environmental codes. *J. Vocat. Behav.* **1982**, *21*, 231–241. [[CrossRef](#)]
60. Gati, I. Description of alternative measures of the concepts of vocational interest: Crystallization, congruence, and coherence. *J. Vocat. Behav.* **1985**, *27*, 37–55. [[CrossRef](#)]
61. Hamming, R. Error detecting and error correcting codes. *Bell Syst. Tech. J.* **1950**, *29*, 147–160. [[CrossRef](#)]
62. Levenshtein, V.I. Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Sov. Phys. Dokl.* **1966**, *10*, 707–710.

63. Grotevant, H.D.; Cooper, C.R.; Kramer, K. Exploration as a predictor of congruence in adolescents' career choices. *J. Vocat. Behav.* **1986**, *29*, 201–215. [[CrossRef](#)]
64. Holland, J.L. *Making Vocational Choices: A Theory of Vocational Personalities and Work Environments*; Prentice-Hall: Englewood Cliffs, NJ, USA, 1985.
65. Strahan, R.F. Measures of consistency for Holland-type codes. *J. Vocat. Behav.* **1987**, *31*, 37–44. [[CrossRef](#)]
66. Frantz, T.T.; Walsh, E.P. Exploration of Holland's theory of vocational choice in graduate school environments. *J. Vocat. Behav.* **1972**, *2*, 223–232. [[CrossRef](#)]
67. Spokane, A.R.; Walsh, W.B. Occupational level and Holland's theory for employed men and women. *J. Vocat. Behav.* **1978**, *12*, 145–154. [[CrossRef](#)]
68. Iachan, R. A family of differentiation indices. *Psychometrika* **1984**, *49*, 217–222. [[CrossRef](#)]
69. Peiser, C.; Meir, E.I. Congruency, consistency, and differentiation of vocational interests as predictors of vocational satisfaction and preference stability. *J. Vocat. Behav.* **1978**, *12*, 270–278. [[CrossRef](#)]
70. Wiggins, J.D.; Moody, A. A field-based comparison of four career-exploration approaches. *Vocat. Guid. Q.* **1981**, *30*, 15–20. [[CrossRef](#)]
71. Abbott, A. Sequence Analysis: New Methods for Old Ideas. *Annu. Rev. Sociol.* **1995**, *21*, 93–113. [[CrossRef](#)]
72. Heine, J.H.; Leitgöb, H.; Hartmann, F.G.; Tarnai, C. Sequenzbasierte Distanzmaße für Kongruenz im Holland-Modell der Interessen- und Umweltorientierungen. In Proceedings of the 50 Kongress der Deutschen Gesellschaft für Psychologie (DGPs), Leipzig, Germany, 18–22 September 2016.
73. Needleman, S.B.; Wunsch, C.D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **1970**, *48*, 443–453. [[CrossRef](#)]
74. Meir, E.I.; Esformes, Y.; Friedland, N. Congruence and differentiation as predictors of workers' occupational stability and job performance. *J. Career Assess.* **1994**, *2*, 40–54. [[CrossRef](#)]
75. Wilkins, K.G.; Tracey, T.J.G. Person environment fit and vocational outcomes. In *Psycho-Social Career Meta-Capacities: Dynamics of Contemporary Career Development*; Coetzee, M., Ed.; Springer International Publishing: Cham, Switzerland, 2014; pp. 123–138.