Journal of Vocational Behavior 136 (2022) 103748

Contents lists available at ScienceDirect

Journal of Vocational Behavior

journal homepage: www.elsevier.com/locate/jvb

Regressed person-environment interest fit: Validating polynomial regression for a specific environment

Stijn Schelfhout [a,b,e,*], Mona Bassleer [b], Bart Wille [c,e], Sofie Van Cauwenberghe [b], Merel Dutry [b], Lot Fonteyne [d], Nicolas Dirix [b], Eva Derous [a,e], Filip De Fruyt [c,e], Wouter Duyck [b,e,f]

[a] Ghent University, Faculty of Psychology and Educational Sciences, Department of Work, Organization and Society, Henri Dunantlaan 2, 9000 Ghent, Belgium

[b] Ghent University, Faculty of Psychology and Educational Sciences, Department of Experimental Psychology, Henri Dunantlaan 2, 9000 Ghent, Belgium

[c] Ghent University, Faculty of Psychology and Educational Sciences, Department of Developmental, Personality and Social Psychology, Henri Dunantlaan 2, 9000 Ghent, Belgium

[d] Ghent University, Department of Educational Policy, Student Counseling Office, Sint-Pietersnieuwstraat 33, 9000 Ghent, Belgium

[e] Ghent University, Interdepartmental Research Group: Vocational and Personnel Psychology (VoPP), Henri Dunantlaan 2, 9000 Ghent, Belgium

[f] The Accreditation Organisation of the Netherlands and Flanders (NVAO), Parkstraat 83, 2514 JG The Hague, the Netherlands

A R T I C L E I N F O

Keywords:
Polynomial regression
PE interest fit
Academic achievement
RIASEC
Vocational interests

A B S T R A C T

Polynomial regression is a proven method to calculate person-environment (PE) interest fit between the RIASEC (realistic, investigative, artistic, social, enterprising and conventional) interests of a student and the RIASEC profile of a study program. The method has shown much larger effects of PE interest fit on academic achievement than earlier approaches in literature. However, the polynomial regression method in its current form only focuses on establishing the regressed interest fit (RIF) of a population of students with their study environments, in order to observe how large the general impact of PE interest fit can become on academic achievement. The present study ($N = 4407$ across $n = 22$ study programs) further validates this method towards new applications by theoretically deriving two measures of RIF that only affect a single environment like a study program. Analyses show that the use of RIF for a single study environment results in an even stronger positive relation between PE interest fit and academic achievement of $r = 0.36$, compared to $r = 0.25$ for the original polynomial regression method. Analyses also show that RIF for one environment can be used to generate interpretable and reliable RIASEC environment profiles. In sum, RIF for a single (study) environment is a promising operationalization of PE interest fit which facilitates both empirical research as well as the practical application of interest fit in counseling settings.

* Corresponding author at: Ghent University, Faculty of Psychology and Educational Sciences, Department of Work, Organization and Society, Henri Dunantlaan 2, 9000 Ghent, Belgium.

E-mail addresses: stijn.schelfhout@ugent.be (S. Schelfhout), mona.bassleer@ugent.be (M. Bassleer), bart.wille@ugent.be (B. Wille), sofie.vancauwenberghe@ugent.be (S. Van Cauwenberghe), merel.dutry@ugent.be (M. Dutry), lot.fonteyne@ugent.be (L. Fonteyne), nicolas.dirix@ugent.be (N. Dirix), eva.derous@ugent.be (E. Derous), filip.defruyt@ugent.be (F. De Fruyt), wouter.duyck@ugent.be (W. Duyck).

https://doi.org/10.1016/j.jvb.2022.103748
Received 1 July 2021; Received in revised form 25 May 2022; Accepted 8 June 2022
Available online 13 June 2022
0001-8791/© 2022 Elsevier Inc. All rights reserved.

1. Introduction

The theoretical notion that people are attracted to, and perform better in environments that fit with their personal characteristics is a contemporary cornerstone of organizational psychology (Nye, Perlus, & Rounds, 2018; van Vianen, 2018), with ample applications in both work (Nye et al., 2012) and higher education (Schelfhout et al., 2019; Schelfhout, Wille, Fonteyne, Roels, De Fruyt, et al., 2021). In his original seminal framework, Schneider (1987) formulated three mechanisms that describe how people (1) are *attracted* to an environment to achieve fit, (2) are *selected* into an environment because of (perceived) fit and (3) possibly also *leave* an environment in case of experienced misfit (i.e., attrition). As a result, the attraction-selection-attrition (ASA) processes imply that the environment is made by the people in that environment (De Cooman et al., 2009; Oh et al., 2018; Schneider, 1987; Schneider et al., 2000). In a similar research line, Holland's (1997) theory proposes that people are inherently motivated to select environments that fit with their personal interests. Moreover, the theory also predicts that individuals working or studying in a fitting environment have a higher chance of obtaining success in their work (Hoff et al., 2020; Hoff et al., 2021) or study environment (Lent et al., 1994; Tracey et al., 2012).

Over the past decades, a vivid and still ongoing line of research has aimed to empirically test the basic tenets of these person-environment (PE) fit theories. The results reported in literature have long remained ambiguous, providing mixed evidence for the assumption that PE interest fit indeed robustly and uniquely predicts individual performance. As an example to illustrate the higher education context of the present study, the reported effects of interest fit on academic achievement ranged between non-existent to very optimistic estimates of about $r = 0.30$ (Nye et al., 2012).

As an explanation for this mixed evidence, van Vianen (2018) stated that the lack of clear findings may originate from issues inherent to existing PE interest fit measures. For instance, Edwards (1994) showed that the vast majority of fit measures put too much constraints on the data, making an unambiguous interpretation of such measures very difficult. van Vianen (2018) also stated that the vocational field was thus in dire need of alternative methodology and new research questions. A recently suggested polynomial regression methodology seems like a promising alternative to calculate PE interest fit, as the methodology renders PE interest fit results that are highly predictive towards work and higher education outcomes (Nye, Butt, et al., 2018; Nye, Prasad, et al., 2018). Polynomial regression calculates PE interest fit by regressing an outcome like academic achievement on the vocational interests of individuals (Edwards, 1993).

With the introduction of a powerful method also comes the necessity to reapply this new method to all areas where vocational interests can make a difference. For instance, polynomial regression could prove crucial as an additional tool to improve academic achievement in higher education. Indeed, internationally, students seem to struggle to make an appropriate and attainable study choice as fail rates in the first year of higher education can become quite high, with estimates ranging from 30 % to even over 60 % (OECD, 2017; Schelfhout et al., 2022).

Assessing PE interest fit is a well-established practice in study orientation (Nauta, 2010) and such assessment can assist in remedying these high rates of failure through study counseling (Schelfhout, Wille, Fonteyne, Roels, De Fruyt, et al., 2021). For instance, students that have to choose a specific study program (or major) in higher education would benefit to know if a specific program would fit their vocational interests. The polynomial regression method in its current form is tailored towards observing how much of the variance in academic achievement of students can be explained by the PE interest fit with their program in the general student population. Considering the importance towards study counseling, the question remains if we can also validate polynomial regression as a way to calculate PE interest fit between a single student and a specific study program.

The present study therefore focuses on determining the PE interest fit of an individual student with one specific (future) study program using polynomial regression. The present study thus theoretically derives a regression from the original polynomial regression formula for PE interest fit that only affects a single study program (Nye, Prasad, et al., 2018). For instance, by regressing an outcome like a GPA on the interests of psychology students only, a typical psychology interest profile is extracted from the data that functions as an environment profile for the psychology program. Such an environment profile can then be used to calculate a PE interest fit between any student and any study program by inserting the corresponding interest parameters of the student (i.e., interest dimension values) and the environment (i.e., interest dimension coefficients) into the regression formula and calculating the regressed value.

1.1. Person environment interest fit

The congruence or fit between person and environment is one of the basic tenets of vocational interests literature (Guan et al., 2021). To ensure that PE interest fit is suitable to aid in study orientation, literature builds on the commensurate interest measurement of both individuals and their work or study environment (Xu & Li, 2020). For instance, Holland's hexagon of vocational interests describes both individuals as well as their working or study environments on six RIASEC (i.e., Realistic, Investigative, Artistic, Social, Enterprising and Conventional) interest dimensions (Holland, 1997). The Realistic dimension mostly involves working with things, like tools and materials. The Investigative dimension involves systematic observation of data, and deduction of information. The Artistic dimension involves creativity, individual expression, and free and unstructured activities. The Social dimension involves working with people, to help them, take care of them, and/or teach or otherwise inform them. The Enterprising dimension involves business-like activities such as leadership and sales, with a strong focus on persuasion and initiative. Finally, the Conventional dimension involves performing highly structured activities and working with rules, procedures, regulations and legislation. The RIASEC profile of an individual can be measured by filling out a RIASEC interest questionnaire, resulting in a score on each of the six interest dimensions. The profile of an environment can be obtained in a number of ways. One way to obtain such a profile uses individuals from that environment as representatives or incumbents of that environment, as was already theorized by Schneider (1987). For a more detailed overview, we refer to Allen and Robbins (2010). Another way to obtain these profiles uses (a combination of incumbent scores and) expert ratings like O*NET (Rounds et al., 1999). If the scores of an individual on the six RIASEC dimensions show certain similarities with the scores of the environment on these six dimensions, it is said that the vocational interests of an individual fit the environment to various degrees.

However, literature has known a long-standing debate on the nature of these similarities. Early high-point coding attempts at determining PE fit only compared the highest one or two scoring dimensions (Young et al., 1998). For instance, in a highly social environment like a nursing program, a student with a dominant social dimension (i.e., with a higher score on the social dimension compared to the other RIASEC dimensions) is considered to have a good PE interest fit with the environment. Although very user-friendly, these methods were eventually heavily criticized, as the criterion for a good fit is almost always inherent to a specific method, making a comparison between methods very difficult at best (Brown & Gore, 1994; Camp & Chartrand, 1992; Tinsley, 2000).

Today, literature still shows a plethora of PE interest fit measures, with various levels of complexity. For instance, correlation fit and Euclidean distance are two continuous and more recent measures that have had some merit in charting the beneficial effects of PE interest fit (Tracey et al., 2012). Correlation fit is a technique that does not compare the highest scores between person and environment, but relies on the scoring pattern of the six dimensions (Schelfhout, Wille, Fonteyne, Roels, De Fruyt, et al., 2021). Calculating the correlation between the RIASEC scores of the person and the RIASEC scores of the environment compares both profile patterns and thus provides a continuous measure of PE interest fit. In contrast, Euclidean distance is a distance-based measure that reduces the profiles of the person and the environment to two points in Euclidean space (Wille et al., 2014). By calculating the distance between these point projections, a continuous measure of PE interest fit is obtained. Literature does report some, albeit small, positive effects (about $r = 0.10$) for both measures on a number of outcomes in work and educational contexts like study choice (Su & Rounds, 2015) and study achievement (Tracey et al., 2012). However, what is considered a good fit remains a challenging question as finding an absolute criterion (e.g., “How well exactly must a student fit a study program before you would advise it as a possible study choice?”) is not that straightforward. Polynomial regression attempts to answer what should be considered a good fit.

1.2. Polynomial regression

Edwards (1993) criticized the construction of PE fit measures (not only concerning vocational interests) in organizational research. In doing so, he also addressed the problem of an objective criterion to distinguish a good fit from a bad one (Edwards, 1994; Edwards & Parry, 1993), advocating that this criterion should be determined in function of the prediction goal. For instance, to investigate the predictive effects of vocational interest on academic achievement, academic achievement should be used as the criterion. In other words, a good fitting student interest profile for a specific study program is a profile that is associated with high academic achievement in that program. Besides the methodological arguments of Edwards (1993), ASA theory also indicates that persons are selected (or retained) into an environment based on perceived fit. As academic achievement already functions as the selection criterion in higher education towards degree attainment (Schneider & Preckel, 2017), academic achievement is not only a methodologically valid criterion choice for PE interest fit but also a theoretically valid one.

In order to practically calculate PE interest fit in function of an objective criterion, Edwards (1994) therefore recommended using polynomial regression. A polynomial is a mathematical expression containing different variables with coefficients, while the expression only includes operations of addition, subtraction, multiplication and non-negative exponentiation (Barbeau, 2003). Polynomial regression as a measure of PE interest fit thus regresses a dependent variable (i.e., academic achievement) on a set of independent variables (i.e., RIASEC interest scores with linear and quadratic terms). As a major advantage for establishing PE interest fit, this regressed interest fit (RIF) does not impose as much constraints on the data as some other methods. Amongst other issues, Edwards (1994) correctly argues that most fit measures (i.e., Euclidean distance) do not represent the components (i.e., PE RIASEC scores) equally, unless one can show that all components have the same variance. Otherwise, the interpretation of such a measure will vary from sample to sample, which is detrimental to the idea of an objective criterion.

In response to these (and other) concerns, Nye, Butt, et al. (2018) introduced polynomial regression as an alternative way to study the effects of PE interest fit on various educational and work outcomes. Applied to RIASEC interest fit, this procedure regresses the criterion of interest (e.g., GPA) on the RIASEC scores of the individual, the RIASEC scores of the environment (e.g., study program), their respective quadratic equivalents and their interactions, which sums to thirty regression terms. More formally, this type of analyses is specified as follows,

$$GPA \sim \sum_{X=R}^{C} \beta X + \sum_{Y=R}^{C} \beta Y + \sum_{X=R}^{C} \beta X^2 + \sum_{Y=R}^{C} \beta Y^2 + \sum_{X,Y=R}^{C} \beta XY + e \tag{1}$$

with X and $X2$ representing the individual linear and quadratic RIASEC terms varying from R to C, Y and $Y2$ representing the environment linear and quadratic RIASEC terms varying from R to C, XY representing the RIASEC interaction terms between individual and environment varying from R to C, and e representing the error term. The (thirty) regression coefficients (β) for each term need to be estimated. In practice, PE interest fit is then calculated by multiplying the student and program RIASEC scores with the corresponding regression coefficients and summing the products. For the present study, we will refer to this general procedure as the RIF30 method, because the polynomial regression to calculate PE interest fit is determined by thirty RIASEC terms. For a further and more recent in-depth description of the RIF30 measure, we refer to Wiegand et al. (2021).

It is important to note that all RIASEC dimensions should remain a part of the regression, even if their direct effect is not significant. Indeed, when conducting regression analyses, one should always consider the possibility of the omitted variable problem (Sackett et al., 2003). When variables related to both the dependent variable and other variables in the model are excluded, the assumption that all relevant predictors are present in the model is violated. This violation can have severe consequences as the estimates of the regression coefficients for the other independent variables can become inaccurate. In the case of vocational interests, we already know from literature that all RIASEC dimensions are correlated to the extent the dimensions exhibit a circular structure (Holland, 1997; Tracey & Rounds, 1996).

Early results indicated that the RIF30 correlated $r = 0.31$ with major GPA, while also providing incremental validity above and beyond the effects of various cognitive and non-cognitive predictors (Nye, Butt, et al., 2018). RIF30 was also correlated up to $r = 0.23$ with work satisfaction and up to $r = 0.34$ with course performance (Nye, Prasad, et al., 2018). These studies show that the power and benefits of this polynomial regression method cannot be underestimated as the method sheds a new light on the importance of vocational interests in both higher education as well as the work field. Despite these benefits, van Vianen (2018) does advocate caution towards three limitations of the polynomial regression method.

As a first limitation, the polynomial regression method requires larger data samples (and environments), depending on the number of predictors (Edwards, 1993; van Vianen, 2018). As an example, Nye and colleagues suggested sample sizes of $N = 1449$ to $N = 30{,}384$, across $n = 74$ to $n = 409$ unique environments respectively (Nye, Butt, et al., 2018; Nye, Prasad, et al., 2018). These examples indicate that on average, each environment is represented by 20 to 74 individuals. For a further discussion on regression sample sizes, we refer to Maxwell et al. (2008). The present study addresses the need for large data samples by showing that a reduced form of polynomial regression can also work for one specific study program environment that has a student population of as low as $n = 65$ students (i.e., within the range mentioned above). However, to ensure that such smaller data samples render valid results and remain representative for the current student population, the present study includes student data of $N = 4407$ first-year students divided over $n = 22$ specific study programs with a wide range of topics, across ten faculties of a large Belgian university (see also Appendix, Table A1).

As a second limitation, polynomial regression (Type I) errors increase with an increasing number of predictors in the regression (Su et al., 2019). The present study addresses this issue by reducing the number of predictors in the regression. Indeed, RIF30 regresses outcomes like achievement on 30 RIASEC terms (see also Eq. (1)): twelve individual terms, twelve environment terms and six interaction terms. The present study reduces the number of predictors needed to calculate RIF to the twelve individual terms.

As a third and final limitation, polynomial regression tends to result in multicollinearity, even when applying counter measures like variable centering (Tinsley, 2000). Multicollinearity is a regression phenomenon in which at least one independent variable is highly correlated with at least one of the other independent variables (Johnston et al., 2018). Although multicollinearity does not have an effect on the explanative power of a regression model, severe multicollinearity can prove problematic as the estimates of the independent variables' coefficients can have very high standard errors, making the coefficient estimates unstable (Kraha et al., 2012). As an example, exclusion of only a few data points can already cause a huge change in the estimates and even the signs of the coefficients of the polynomial regression. In practice, a good PE interest fit between student and program will not change due to multicollinearity, but multicollinearity does make it very hard to assess *why* the student shows such a good fit with the program (i.e., which RIASEC dimensions are responsible). The present study addresses this issue by showing that RIF for one specific program does not suffer from multicollinearity if the RIF is limited to six RIASEC linear terms only.

1.3. RIF12 for one specific study program environment

A RIF between one student and one study program is needed to further validate polynomial regression as a powerful and practically useful way of calculating PE interest fit. Specifically, it can be argued that the RIF30 method in its current form is less suited to calculate PE interest fit between one individual (i.e., student) and a specific environment (i.e., study program) as the regression also contains the data from other environments. Indeed, the obtained regression coefficients for RIF30 can vary depending on the sample, as was also addressed in earlier polynomial regression studies on PE interest fit (Nye, Butt, et al., 2018; Nye, Prasad, et al., 2018). This variation could prove problematic for reliably calculating the fit of a student with a study program. For example, a regression based on student data from the study programs of psychology, law and economics can have totally different regression coefficients compared to a regression based on student data from the programs of psychology, mathematics and medicine. However, when calculating the PE interest fit of a psychology student with the psychology program, the calculation could render different results depending on the data set used, while PE interest fit is still calculated between the same student and the same study program environment. Obtaining different results using the same method when calculating PE interest fit between a student and a study program would be detrimental to the concept of PE interest fit.

Van Iddekinge et al. (2011) have touched upon polynomial regression to calculate the interest fit between an individual and a specific environment in their meta-analytic study on the effects of vocational interest on employee performance and turnover. Although this meta-analysis was unable to perform a full regression due to the absence of raw data, the analysis did manage to show that even meta-analytic compositions of vocational interests across studies (e.g., scores on all six RIASEC scales) had a clear edge over classic congruence indices of PE interest fit and single RIASEC dimensions. Indeed, these composites revealed better criterion validity on various outcomes, while also clearly distinguishing how large the effects of individual interest dimensions were when controlling for the other dimensions. Important to note, Van Iddekinge et al. (2011) analyzed these composite measures in separate, specific environments of workers that executed similar jobs. In other words, the authors did not ignore the environment, but delineated a specific environment based on the professional activity of the individuals working in that environment. For instance, individuals that were occupied as an accountant were considered to work in the environment of accountancy. This specific delineation of environments is in line with vocational interest theory as several studies have indicated that a specific environment attracts individuals with similar interest profiles (Nye, Perlus, & Rounds, 2018; Schelfhout et al., 2019), as predicted by theory (Holland, 1997; Schneider, 1987).

The present study wants to delineate specific environments in higher education. For instance, a psychology student is considered to study in the environment of a psychology program. As such, every psychology student in that program shares the specific psychology study program environment, with the same study content for each student (e.g., courses and lectors).[1] For example, if one uses the RIASEC scores of the psychology students as incumbents for the psychology program, averaging out the RIASEC scores would render a RIASEC environment profile for the psychology program that is equal for all psychology students. If this psychology student population is used exclusively (i.e., no students from other programs) in Eq. (1), all eighteen environmental terms from Eq. (1) become constants (i.e., twelve environment terms) or redundant multiples of the individual terms (i.e., six interaction terms) as the psychology program environment has the same properties for each student. In a regression, constants and redundant multiples should be eliminated altogether (i.e., their regression coefficients should be put to zero), as their purpose is nil. In other words, within one specific environment, PE interest fit through polynomial regression can be statistically reduced to twelve individual linear and quadratic RIASEC terms,

$$GPA \sim \sum_{X=R}^{C} \beta X + \sum_{X=R}^{C} \beta X^2 + e \tag{2}$$

with X and $X2$ representing the individual linear and quadratic RIASEC terms varying from R to C and e representing the error term. The regression weight (β) for each term needs to be estimated. In practice, PE interest fit for a student is calculated by multiplying the student RIASEC scores with the corresponding program regression coefficients and summing the products. Although Eq. (2) no longer features environmental parameters, it is important to note that the environment is still taken into account as the RIASEC (and GPA) scores of the individual student are compared to the (regressed) scores of its peers from the same environment. For the present study, we will call this method of calculating RIF the RIF12 method, as it takes twelve terms to perform the regression.

The relation between RIF30 and RIF12 is determined by the number of specific environments m used in RIF30. In fact, if there is only one environment such as a specific study program environment (i.e., one set of environment parameters that are identical for all individuals like in the example given above), RIF30 and RIF12 are identical. More formally,

$$\begin{aligned} GPA &\sim \left(\sum_{X=R}^{C} \beta X + \sum_{Y=R}^{C} \beta Y + \sum_{X=R}^{C} \beta X^2 + \sum_{Y=R}^{C} \beta Y^2 + \sum_{X,Y=R}^{C} \beta XY + e \mid \mathrm{m} = 1 \right) \\ &= GPA \sim \sum_{X=R}^{C} \beta X + \sum_{X=R}^{C} \beta X^2 + e \end{aligned} \tag{3}$$

with X and $X2$ representing the individual linear and quadratic RIASEC terms varying from R to C, Y and $Y2$ representing the environment linear and quadratic RIASEC terms varying from R to C, XY representing the RIASEC interaction terms between individual and environment varying from R to C, m representing the number of unique environments and e representing the error term. The regression weight (β) for each term needs to be estimated.

As a result of the reduction of terms, RIF12 for one specific program requires 60 % (12 vs. 30) fewer estimated coefficients compared to RIF30, limiting the effects of estimation errors on PE interest fit. Predicting academic achievement can thus be done more precisely in one specific program, as the parameters from the other programs do not have to be considered.

H1. RIF12 generates a more accurate PE interest fit due to a smaller estimation error compared to RIF30.

By definition, a practical application of RIF12 to one discipline or program environment no longer harbors between-discipline parameters. However, it is important to note that RIF12 as a method does not ignore the between-environment variance. On the contrary, by applying RIF12 to each of the programs specifically and separately, the variance present in both RIASEC measures and the GPA scores is split up across study programs. In a way, RIF12 distributes the RIASEC and GPA variance in the population by controlling for the differences between study programs. Indeed, literature agrees that different programs will reward different student interest RIASEC patterns (Smart et al., 2000). By allowing the RIF12 regression coefficients to vary between-disciplines (i.e. different regression coefficients) in accordance to the patterns present in the specific disciplines (i.e., which RIASEC profiles lead to better GPA), the relation between GPA and PE fit will become stronger compared to RIF30. As a consequence, the regression of GPA on the RIASEC variables will lead to a higher explained variance compared to RIF30, which does not feature such a between-program control.

H2. RIF12 generates a PE interest fit estimate that has a larger effect on academic achievement compared to RIF30.

Classic congruence indices are less powerful in detecting the positive relation between PE interest fit and academic achievement (Nye, Prasad, et al., 2018). Up until now however, studies have not yet compared polynomial regression to continuous methods of establishing PE fit like Euclidean distance or correlation fit. Still, the same rationale regarding the classic, more categorical indices can be applied to the more recent, continuous ones (Edwards, 1993). Indeed, both types of methods put too much constraints on the data, artificially limiting the common variance between criterion and regressors, which renders a less powerful method that underestimates the effects of PE interest fit. We thus predict that:

[1] As noted by one reviewer, we do acknowledge that a student's environment is not limited to the study environment and that not all students experience the (study) environment in a similar way.

H3. RIF generates PE interest fit estimates that have a larger effect on academic achievement compared to Euclidean distance and correlation fit.

1.4. Study environment profile generation using RIF6

The current methods of determining environment profiles that are needed to subsequently establish PE interest fit measures are still suboptimal (Nauta, 2010). Despite the empirical evidence on the importance of vocational interests variance within one specific environment (Nye, Perlus, & Rounds, 2018;Schelfhout et al., 2019; Tracey et al., 2012), the most common methods of determining RIASEC environment profiles use RIASEC interest means, while ignoring RIASEC interest variance. For instance, the incumbent method usually only takes the mean of RIASEC scores over students in a specific program to determine the RIASEC profile of that specific program (Allen & Robbins, 2010; Fonteyne et al., 2017; Schelfhout, Wille, Fonteyne, Roels, De Fruyt, et al., 2021). Indeed, even a specific, homogeneous environment (for instance due to ASA processes, see also Schneider et al., 2000) can still exhibit quite some variance regarding vocational interests, which is rarely considered when determining the vocational interest profile of a specific environment (Nye, Perlus, & Rounds, 2018). Such environment interest variance can even affect important outcomes like academic achievement (Schelfhout et al., 2019; Tracey et al., 2012).

As polynomial regression explicitly uses unconstrained vocational interest variance to explain outcomes like academic achievement, such a regression seems an excellent candidate to generate environment profiles more accurately. However, RIF30 does not use data on one study program, but uses data on a multitude of programs that can have totally different interest patterns (Smart et al., 2000). RIF12 for one specific study program solves this issue as RIF12 estimates regression coefficients for all RIASEC dimensions based on the RIASEC scores of students in that specific program. As such, the variance of these RIASEC dimensions over students in that specific program is also taken into account. As a consequence, students that have a RIASEC profile that fits the (regression) coefficients from their program profile, will have a high PE interest fit value and are thus theoretically predicted to obtain good results in that program (Holland, 1997). Although RIF12 seems a promising method of estimating a specific RIASEC environment profile, a RIF12 profile would still incorporate quadratic RIASEC terms, which are by definition collinear iterations of linear RIASEC terms. Multicollinearity can become a serious issue if the regression models and their coefficients are used for other purposes than explanation only. The multicollinearity problem can distort the estimates of the coefficients, making the environment profiles unreliable (Kraha et al., 2012). However, literature already shows that linear RIASEC dimensions are related in a circular pattern, but not to the extent that two or more dimensions have highly collinear patterns (Fonteyne et al., 2017; Holland, 1997). For profile generation specifically, we thus consider the option to further reduce RIF12 to a regression with linear RIASEC terms only,

$$GPA \sim \sum_{X=R}^{C} \beta X + e \tag{4}$$

with X representing the individual linear RIASEC terms varying from R to C and e representing the error term. The regression weight (β) for each term needs to be estimated. The six-term RIASEC expression is still considered a polynomial, albeit of the first degree. In practice, PE interest fit for a student is again calculated by multiplying the student RIASEC scores with the corresponding program regression coefficients and summing the products. For the present study, we will call this method of calculating RIF the RIF6 method, as it takes six terms to perform the regression. Note that RIF6 and RIF12 do not require an estimate of the RIASEC environment scores prior to the analyses as there are no longer any environment profile elements present in the formula. In contrast to many other incumbent methods, these environment scores (i.e., regression coefficients) are part of the results.

By removing the quadratic terms and by using the variance of RIASEC scores in the student incumbents, we expect that RIF6 is a valid alternative for environment RIASEC profile generation as was requested in literature (Nauta, 2010; Nye, Perlus, & Rounds, 2018). As such, we predict that

H4. RIF6 does not cause multicollinearity issues.

When describing the proposed regressions (i.e., RIF12 and RIF6), the present study also considers the statistical consequences of omitting predictors (i.e., the quadratic terms in reducing RIF12 to RIF6) in these regressions. In case of such an omission, the explained variance of the omitted variables is absorbed into the error term (see also Eqs. (1), (2), (3) and (4)), which still makes for a balanced regression. Yet, one should be aware of the consequences of such an operation. In case the omitted variables were initially explanatory towards the dependent variable, the error term will increase in importance (i.e., unexplained variance), while the amount of variance explained by the independent variables will drop. If, for any reason, such an omission is conducted, the cost of this omission should be calculated and weighed against the advantages of having a reduced model. For calculating the cost, the present study therefore provides an estimate of how much explanative power is lost towards academic achievement when using RIF6 over RIF12.

2. Method and materials

2.1. Data

The data were drawn from a large regionwide project (Flanders, Belgium) providing study counseling advice towards high school students who are making the transition into higher education in an open access education system. The project received favorable advice of the responsible ethics committee at Ghent University (department of Experimental Psychology, application number *2016/*

Table 1
Variable summary and bivariate correlations.

	M	*SD*	GPA	GRSE	COR	ED	RIF30	RIF12	R	I	A	S	E	C
GPA	445.31	242.59	1	0.36**	0.08**	−0.04*	0.25**	0.36**	0.00	0.04**	−0.02	0.02	0.01	−0.01
GRSE	72.02	6.53		1	0.02	0.02	0.11**	0.17**	0.02	0.15**	0.03	0.01	−0.06**	−0.07**
COR	0.71	0.28			1	−0.43**	0.32**	0.26**	−0.26**	−0.15**	−0.18**	0.07**	0.00	−0.10**
ED	92.19	51.12				1	−0.12**	−0.10**	0.19**	0.20**	0.45**	0.16**	0.07**	0.09**
RIF30	445.35	60.42					1	0.67**	0.03	0.17**	−0.06**	0.09**	0.04*	−0.04**
RIF12	446.94	87.80						1	0.00	0.11**	−0.04*	0.05**	0.01	−0.05**
R	19.70	24.75							1	0.39**	0.14**	−0.22**	0.16**	0.26**
I	33.32	21.00								1	0.21**	0.20**	−0.01	0.12**
A	28.72	25.07									1	0.43**	0.26**	0.08**
S	34.32	26.39										1	0.15**	0.06**
E	35.63	28.57											1	0.73**
C	23.20	24.09												1

Note. GPA = grade point average, GRSE = global result secondary education, COR = correlation fit, ED = Euclidean distance, RIF30 = regressed interest fit using full polynomial regression with 30 terms, RIF12 = regressed interest fit using full polynomial regression with 12 terms, R = realistic interest score, I = investigative interest score, A = artistic interest score, S = social interest score, E = enterprising interest score, C = conventional interest score. Students' RIASEC interest scores are obtained through the SIMON-I interest inventory.

* $p < .05$.

** $p < .01$.

82). The education system in which the project is embedded is characterized by a relatively low entry cost (i.e., a maximal tuition fee of about € 1000 or $ 1170) and an open access policy (i.e., no admission tests or entry level GPA requirements). Such an environment also provides an excellent opportunity to investigate the influence of vocational interests on study choice and academic achievement (Schelfhout, Wille, Fonteyne, Roels, Derous, et al., 2021), in the absence of additional (financial and achievement-related) requirements.

For the present prospective study, the dataset features interest and performance data from a large overall sample of $N = 4407$ first year generation (i.e., first registration) students (59 % female, with a mean age of $M = 18.16$, $SD = 0.82$ and a median age of $Mdn = 18$, which covered 88 % of the sample), distributed across ten faculties and 22 bachelor programs of a Belgian university that is ranked in the top 100 of the Academic Ranking of World Universities (ARWU, formerly known as the Shanghai Ranking, see also https://www.shanghairanking.com/rankings).

At the start the academic year 2018–2019 first-year students were strongly advised through multiple channels (e.g., lectors, emails and messages distributed through online student platforms) to fill out an online RIASEC questionnaire, specifically designed for the transition towards higher education (Fonteyne et al., 2017). At the end of the academic year, exam data about their global grade point average (GPA) for their complete program were linked to their interest data. Out of a possible 5699 registered students for the $n = 22$ programs, 4479 students completed the RIASEC survey (response rate = 79 %). Also, out of these possible 5699 registered students, 571 dropped out before taking exams (dropout = 10 %). Combining both data collections (i.e., RIASEC questionnaires and exam GPA), a total of $N = 4,407$ students was retained of which we had data on both RIASEC as well as GPA, which is about 77 % of the total number of registered students. About 25 % of all students originated from a lower SES background (i.e., lower family income and/or lower parental literacy).

It is important to note that within each study program all generation students on the model trajectory have identical curricula. Such curricula comprise a number of courses which, together, provide a delineated study environment for each specific program. Though some programs have somewhat overlapping topics (e.g., Biomedical Sciences and Biochemistry and Biotechnology), the study environment for each program is quite unique as each program has a combination of different courses, lectured by different professors. Each course is graded after an exam. For more information about the study choice process and the evaluation procedures used in this educational system, we refer to Schelfhout et al. (2022).

2.2. *Measures*

Table 1 shows the variable summary and the correlation matrix of all variables. Variables were centered prior to RIF analyses to proactively counter multicollinearity.

2.2.1. *Academic achievement*

Academic achievement was measured at the end of the academic year as the weighted average score (on a scale from 0 to 1000) of a student across all courses included in his/her specific program. The weights are determined analogous to the ECTS (European Credits Transfer system) credits of each course, as was introduced by the Bologna declaration (1999). GPA measures were linked to students' vocational interests through a central databank at the governing university.

2.2.2. *RIASEC dimensions of vocational interests*

The RIASEC dimensions are measured using the SIMON-I instrument, specifically designed for the context of study counseling in higher education (Fonteyne et al., 2017). The test consists of 173 items with a yes or no answer, asking students whether they would be interested in performing specific activities and occupations. As such, the realistic dimension is measured through 27 items ($\alpha = 0.93$) like “Developing electronic systems?” and “Forrester?”. The investigative dimension is measured through 33 items ($\alpha = 0.88$) like “Analyzing statistics?” and “Biologist?”. The artistic dimension is measured through 30 items ($\alpha = 0.92$) like “Designing webpages?” and “Editor?”. The social dimension is measured through 32 items ($\alpha = 0.92$) like “Giving travel advice?” and “Nurse?”. The enterprising dimension is measured through 26 items ($\alpha = 0.93$) like “Conducting a meeting?” and “Sales manager?”. And finally, the conventional dimension is measured through 25 items ($\alpha = 0.92$) like “Monitor quality standards?” and “Safety advisor?”. The final student score on each dimension is a number between 0 and 100 and is calculated by summing the number of “yes” answers on one dimension, dividing this sum by the total number of items for that dimension and then multiplying this quotient with 100. For instance, if a student answered “yes” to twenty items of the conventional dimension, the student scores 20 out of 25 or 80 (out of 100) for the conventional dimension. A higher dimension score indicates a stronger preference for that interest field.

We tested the hypothesized circular structure of the RIASEC interests. A randomization test of hypothesized order relations (RTOR) showed a correspondence index of $CI = 0.83$, $p = .03$, which is considered a good fit, even when adhering to the most conservative cutoff of $CI > 0.70$ (Rounds & Tracey, 1996).

2.2.3. *Correlation fit*

A Pearson's product-moment correlation coefficient (or bivariate correlation) is obtained for every student by calculating the correlation between the RIASEC profile of the student and the RIASEC profile of the chosen program. The program profile takes the exact form of a student profile with interest scores for each of the six RIASEC dimensions (e.g., R = 10, I = 5, A = 30, S = 80, E = 26, C = 10). For correlation fit and Euclidean distance (see further), the twenty-two program profiles were obtained from a large independent dataset ($N_0 = 4932$, with a response rate of about 63 %) by averaging out the RIASEC scores across senior students in each program at the same university as where the data from N originated. The procedure to calculate program profiles based on this independent dataset was analogous to Allen and Robbins (2010).The senior students showed academic success and perseverance by completing at least the first two years of their program (Schelfhout, Wille, Fonteyne, Roels, De Fruyt, et al., 2021). After the online survey, <3 % of these senior students indicated they would not take the program again if given the opportunity, which is an indication these students were still interested in their program of choice. In order to further test the validity of these program RIASEC measures, we correlated the program RIASEC means of N_0 with the program RIASEC means of N. The mean correlation over programs was $r = 0.95$ with a standard deviation of $SD = 0.07$.

2.2.4. Euclidean distance

Euclidean distance is calculated analogous to Wille et al. (2014). Two points with each two coordinates in Euclidean space are obtained for each student. The people–things (P/T) axis runs from S to R on the RIASEC hexagon. The P/T coordinate for each student and each program is calculated as follows: $\mathrm{P/T} = 2 \times \mathrm{R} + \mathrm{I} - \mathrm{A} - 2 \times \mathrm{S} - \mathrm{E} + \mathrm{C}$. The data–ideas (D/I) axis runs in between the E and C dimension to in between the A and S dimension on the RIASEC hexagon. The D/I coordinate for each student and each program is calculated as follows: $\mathrm{D/I} = 1.73 \times \mathrm{E} + 1.73 \times \mathrm{C} - 1.73 \times \mathrm{I} - 1.73 \times \mathrm{A}$. Finally, Euclidean distance (ED) is calculated for each student as follows: $\mathrm{ED} = \mathrm{SQRT}\ ((\text{student P/T} - \text{program P/T})^2 + (\text{student D/I} - \text{program D/I})^2)$.

2.2.5. RIF30

RIF30 is calculated using Eq. (1). GPA is regressed on the thirty-term polynomial as presented in the introduction. For each term, a regression coefficient is estimated. For each participant, PE interest fit is calculated by summing the thirty products of all person and environment RIASEC term scores and their corresponding regression coefficients. The fit error is the absolute difference between GPA and RIF30.

2.2.6. RIF12

RIF12 is calculated using Eq. (2). GPA is regressed on the twelve-term polynomial as presented in the introduction, for each of the 22 programs separately. For each term, a regression coefficient is estimated. For each participant in a specific program, PE interest fit is calculated by summing the twelve products of all individual RIASEC term scores with the corresponding coefficients from the program-specific regression. The fit error is the absolute difference between GPA and RIF12.

2.2.7. RIF6

RIF6 is calculated using Eq. (4). GPA is regressed on the six-term polynomial as presented in the introduction, for each of the 22 programs separately. For each term, a regression coefficient is estimated. For each participant in a specific program, PE interest fit is calculated by summing the six products of all individual RIASEC term scores with the corresponding coefficients from the program-specific regression. The fit error is the absolute difference between GPA and RIF6.

2.2.8. Global result secondary education

To assess the explanative power of polynomial regression towards academic achievement, we have controlled for the effects of prior achievement, as prior achievement is arguably considered as the best predictor towards future achievement (Schneider & Preckel, 2017). Students thus self-reported their global result in the final year of secondary education (GRSE) through a score ranging from 1 to 100. The GRSE variable acts as a benchmark (i.e., as a control variable) to assess the magnitude of the effects of RIF on academic achievement.

2.3. Analyses

For H1, a paired, two-sample *t*-test is used on the absolute fit errors of student RIF (i.e., the difference between the regressed outcome or RIF and the actual outcome or GPA) to test if RIF12 indeed renders a more accurate fit compared to RIF30. Effect size is indicated using a Cohen's *d* (0.01 – very small effect; 0.20 – small effect, 0.50 – medium effect; 0.80 – large effect; 1.20 – very large effect; 2.00 – huge effect; for an overview, see Sawilowsky, 2009). For H2, a regression of GPA is used to test if RIF12 has a larger effect on academic achievement compared to RIF30. For H3, this GPA regression is also used to test if RIF has a larger effect on academic achievement than correlation fit and ED. GRSE is included as a control variable that also functions as a benchmark for the predictive power of PE interest fit towards academic achievement (Schelfhout et al., 2022; Schneider & Preckel, 2017). For H4, we have examined multicollinearity in all polynomial regressions, by making use of the variance inflation factor (VIF) and the condition index (CI). The VIF quantifies to which extent an independent variable is related to the other independent variables in the regression (Sheather, 2009). Literature has suggested several benchmarks to decide whether VIF values in a regression are too high. The most conservative benchmarks suggest a threshold of $\mathrm{VIF} < 2.50$ to reject multicollinearity altogether (Everitt & Skrondal, 2010; Johnston et al., 2018). If the benchmark of $\mathrm{VIF} < 2.50$ is exceeded, a variance decomposition can reveal which dimensions show collinearity (i.e., high loadings of two or more variables on the same decomposition dimension). Subsequently, the condition index (CI) evaluates if the observed collinearity is considered problematic. Literature has suggested several benchmarks for interpreting CI, with $\mathrm{CI} > 15$ indicating multicollinearity and $\mathrm{CI} > 30$ indicating strong multicollinearity (Jackson, 2017; Kennedy, 2003). We have estimated the cost of using a RIF6 over a RIF12 through the use of a paired two-sample *t*-test.

3. Results

3.1. RIF12 versus RIF30

H1 stated that RIF12 generates a more accurate PE interest fit. The paired sample *t*-test on the average fit errors of RIF30 (M = 197.71, SD = 126.96) and RIF12 (M = 188.19, SD = 126.26) over students is significant, $t\,(4406) = 10.70$, $p < .001$, Cohen's $d = 0.16$. The *t*-test on the average fit errors of RIF30 (M = 201.51, SD = 27.52) and RIF12 over programs (M = 185.00, SD = 28.44) is also significant, $t\,(21) = 4.49$, $p < .001$, Cohen's $d = 0.96$. Considering the results of both tests, H1 is confirmed. RIF12 generates a 5 % (i.e., 197.71 / 188.19 − 1) more accurate PE interest fit over students and a 9 % (i.e., 201.51 / 185.00 − 1) more accurate PE interest fit over programs compared to RIF30.

H2 stated that RIF12 generates a PE interest fit that has a larger effect on academic achievement than PE interest fit generated using RIF30. For H2, GPA is regressed on RIF30, RIF12, correlation fit, Euclidean distance and GRSE. The omnibus test was significant, $F\,(5, 4198) = 237.27$, $p < .001$, $R^2 = 0.22$. RIF30 ($p = .51$). Looking at individual predictors, correlation fit ($p = .67$) and Euclidean distance ($p = .55$) no longer have a significant effect on GPA, while RIF12 ($\beta = 0.29$, $p < .001$) and GRSE ($\beta = 0.31$, $p < .001$) do have a significant (standardized) effect on GPA. As an indication of effect size (without controlling for the other measures, see also Table 1), PE interest fit using RIF30 explains about 6 % (0.25 × 0.25) of the variance in GPA, while PE interest fit calculated using RIF12 and GRSE both explain about 13 % (0.36 × 0.36). Controlling for the effect of GRSE (i.e., $R^2 = 0.13$) exclusively, RIF12 adds a unique explained variance of about 9 % (and vice versa), while RIF30 only adds 4 %. As such, PE interest fit calculated using RIF12 indeed generates a larger and more important effect on academic achievement than RIF30, confirming H2.

H3 stated that RIF generates PE interest fit estimates that have a larger effect on academic achievement than PE interest fit generated using Euclidean distance and correlation fit. Analogous to H2, the regression of GPA on RIF30, RIF12, correlation fit, Euclidean distance and GRSE also shows that PE interest fit calculated using RIF12 has a larger and more important effect on academic achievement than correlation fit and Euclidean distance, confirming H3. As additional evidence, correlation fit and Euclidean distance explained <1 % of the variance in GPA.

3.2. Multicollinearity

For RIF30, the highest VIF = 329.63 already suggested there was evidence of multicollinearity. Closer inspection revealed that 45 % (14 out of 31) of the variance decomposition dimensions had a CI > 15 and 19 % (6 out of 31) dimensions had a CI > 30. As an example, variance decomposition revealed that the highest CI = 90.95 was due to high loadings of the environment terms of E (0.93), C (0.69), E^2 (0.76) and C^2 (0.68) on one specific dimension of the variance decomposition. We therefore conclude that RIF30 is unsuited for environment profile generation as results show too much indications of multicollinearity, making the coefficient estimates unstable.

For RIF12 (see Table 2), all program-specific regressions showed indications of multicollinearity as VIF < 2.50 did not hold for any program-specific regression. Closer inspection of all program regressions showed that CI < 15 was violated in all 22 programs, while CI

Table 2
Variance inflation factor in regressed interest fit models.

Programs	*N*	Highest VIF RIF6	Highest VIF RIF12
Psychology	474	1.94	26.38*
Communication Sciences	106	2.04	29.39*
Educational Sciences	131	2.08	71.78*
Political Sciences	74	2.29	27.88*
Law	330	2.00	23.92*
Criminological Sciences	185	2.22	19.52*
Speech and Hearing Sciences	68	2.03	55.20*
Physical Education	65	2.54*	26.49*
Linguistics and Literature	153	2.05	29.60*
History	85	2.26	23.18*
Physical Therapy and Rehabilitation	297	2.10	22.71*
Pharmaceutical Sciences	241	2.09	22.87*
Bioscience Engineering	217	1.88	29.86*
Economics	443	1.91	22.90*
Biomedical Sciences	218	2.54	33.88*
Engineering - Architecture	75	2.22	31.59*
Engineering	308	2.26	21.89*
Business Economics	345	2.18	23.57*
Bioscience Engineering Technology	74	1.97	20.09*
Engineering Technology	311	2.36	22.93*
Applied Language Studies	137	2.21	22.72*
Biochemistry and Biotechnology	70	5.43*	59.84*

Note. VIF = variance inflation factor with * VIF > 2.50, RIF12 = regressed interest fit using polynomial regression with 12 terms, RIF6 = regressed interest fit using polynomial regression with 6 terms.

< 30 was violated in five programs (23 %). We therefore conclude that RIF12 also shows too much evidence of multicollinearity. As the coefficient estimates are not stable enough, we deem RIF12 unsuited for environment profile generation.

H4 stated that RIF6 does not cause multicollinearity issues. For the RIF6 measure, only two program-specific regressions (9 %) showed a minor indication of multicollinearity due to violating VIF < 2.50. Closer inspection of both regressions (i.e., physical education program and biochemistry and biotechnology program) showed CI maxima of CI $= 4.32$ and CI $= 6.17$, which do not exceed the CI thresholds of CI < 15 or CI < 30. As such, we confirm H4 that RIF6 does not suffer from multicollinearity and thus can be used for profile generation. Table 3 shows the estimated program profiles (standardized regression coefficients) using RIF6 for all programs.

We estimated the cost of using linear regression by testing the error difference between PE interest fit generated by RIF12 ($M = 188.19$, $SD = 126.26$) and RIF6 ($M = 192.04$, $SD = 126.92$). Results showed that the difference was significant, $t\,(4{,}406) = 6.08$, $p < .001$, Cohen's $d = 0.09$. Using RIF6 thus had a 2 % accuracy cost (192.04 / 188.19 − 1), while also having a cost of explanative power as the correlation between academic achievement and PE interest fit dropped from $r = 0.36$, $t\,(4405) = 25.34$, $p < .001$, $R^2 = 0.13$ to $r = 0.32$, $t\,(4405) = 22.06$, $p < .001$, $R^2 = 0.10$.

4. Discussion

Polynomial regression is a method for establishing fit between variables, without putting too much constraints on the data (Edwards, 1993; Edwards, 1994; Edwards & Parry, 1993; van Vianen, 2018). Recent studies on vocational interests indicate that a good regressed interest fit (RIF) between the vocational interests of a student or worker and the chosen study program or job leads to better work and study performance (Nye, Prasad, et al., 2018), vastly outperforming traditional methods like high-point coding (Young et al., 1998), and empirically confirming the theoretical predictions of Holland (1997). However, with these new insights also comes the necessity to disseminate this more powerful method and ensure these beneficial effects of PE interest fit are detectable in all areas where vocational interests can make a difference. For instance, PE interest fit can make a difference in study counseling to guide students towards fitting and attainable programs (Schelfhout, Wille, Fonteyne, Roels, De Fruyt, et al., 2021). Unfortunately, the polynomial regression method still showed three limitations that restrict the further generalization of this method towards applications like study counseling. Polynomial regression (1) requires a large number of participants over a large number of environments (Nye, Butt, et al., 2018; van Vianen, 2018), (2) is prone to error inflation for each term that enters the regression (Su et al., 2019; van Vianen, 2018) and (3) is also prone to multicollinearity (Tinsley, 2000; van Vianen, 2018). By addressing these limitations, the present study thus validates polynomial regression as a method for calculating the PE interest fit of one student to a single study program environment by delineating an environment for each specific program. The performance of RIF12 (i.e., a regression with twelve individual RIASEC scores) and RIF6 (i.e., a regression of six individual RIASEC scores) for single environments were empirically tested against the performance of the original polynomial regression formula RIF30 (i.e., a regression of 30 terms based on the individual and environment RIASEC scores) using the data of $N = 4407$ first-year students divided over 22 specific study programs, across ten faculties of a large Belgian university. As the present study is conducted in an open access and low-cost educational system, student choice of study program can occur freely in absence of GPA – requirements or hefty tuition fees (Schelfhout et al., 2019).

Table 3
RIF6 study program profiles.

Program	R	I	A	S	E	C
Psychology	−0.034	−0.055	0.023	0.116	0.027	−0.073
Communication Sciences	−0.148	0.245	−0.092	−0.161	0.220	−0.303
Educational Sciences	−0.059	0.116	−0.245	−0.037	0.193	−0.092
Political Sciences	−0.092	0.387	−0.299	−0.034	0.283	−0.313
Law	−0.080	0.027	0.065	−0.073	0.151	−0.068
Criminological Sciences	−0.218	0.021	0.067	−0.056	0.151	−0.017
Speech and Hearing Sciences	−0.099	−0.011	0.196	0.021	−0.017	−0.063
Physical Education	−0.194	0.231	0.045	0.240	−0.097	−0.061
Linguistics and Literature	−0.191	0.190	0.047	0.025	0.210	−0.092
History	−0.274	0.363	0.051	0.134	0.022	−0.127
Physical Therapy and Rehabilitation	0.004	0.120	−0.077	0.224	−0.062	−0.116
Pharmaceutical Sciences	−0.149	0.182	−0.104	0.085	−0.009	−0.027
Bioscience Engineering	−0.090	0.097	0.073	−0.029	−0.010	−0.005
Economics	0.005	0.068	−0.050	0.008	0.129	−0.108
Biomedical Sciences	−0.066	0.027	0.152	−0.143	−0.003	0.083
Engineering - Architecture	0.014	−0.109	0.120	−0.206	0.083	0.041
Engineering	−0.090	0.199	0.044	−0.225	0.083	−0.059
Business Economics	−0.164	0.004	0.002	0.018	0.070	0.059
Bioscience Engineering Technology	0.018	0.116	−0.367	0.238	−0.057	0.023
Engineering Technology	−0.062	0.074	0.094	−0.156	−0.083	0.001
Applied Language Studies	−0.071	−0.095	0.247	−0.055	0.017	−0.054
Biochemistry and Biotechnology	0.078	0.064	−0.071	0.061	0.079	−0.247

Note. RIF6 = regressed interest fit with six terms, R = realistic dimension, I = investigative dimension, A = artistic dimension, S = social dimension, E = enterprising dimension, C = conventional dimension. The program profiles are the result of the regression of GPA on the RIASEC scores of students in these programs. The regression coefficients were standardized so that the program profile RIASEC dimensions are comparable within and between programs.

4.1. Empirical findings

The present study thus found that PE interest fit calculated using RIF12 for a specific study program had an even stronger positive effect on academic achievement comparing to RIF30. RIF12 explained about 13 % of the variance in the academic achievement of students, more than doubling the effect of RIF30 of about 6 %. Compared to literature, these results even exceed the most optimistic estimates of about 10 % explained variance in academic achievement (Nye et al., 2012). The effects of RIF12 were about as large as the effects of high school performance, which is considered one of the best predictors of academic achievement (Schneider & Preckel, 2017). These program-specific RIF measures thus upgrade the status of vocational interests as a prime predictor of academic achievement, on par with predictors like previous achievements and cognitive ability (Schelfhout et al., 2022). These results replicate and further strengthen the empirical evidence regarding the power of polynomial regression to calculate PE interest fit. The stronger positive effect of RIF12 (vs. RIF30) on academic achievement is a direct consequence of 5 % smaller estimation errors of polynomial regression in a specific environment, in comparison to polynomial regression with a range of different environments. When comparing both error rates over programs, we even observe a somewhat large effect of about 9 %. We expected these effects for two main reasons. First, RIF12 has to estimate 60 % fewer regression coefficients, which limits the influence of estimation errors. And second, different vocational environments reward different interest patterns (Smart et al., 2000). It is therefore easier to detect the relevant RIASEC pattern in one specific environment compared to an amalgam of different vocational environments. The power of RIF12 as a measure of PE interest fit also proved superior to correlation fit and Euclidean distance, as both measures only showed marginal effects on academic achievement of < 1 % of explained variance in higher education GPA (Schelfhout et al., 2019; Tracey et al., 2012). These effects were also no longer significant when controlling for RIF12. These results are in line with the theoretical predictions from Edwards, as correlation fit and Euclidean distance methods put too many constraints on the data to empower stronger correlations between PE interest fit and academic achievement (Edwards, 1993; Edwards, 1994; Edwards & Parry, 1993; van Vianen, 2018).

Several studies already pointed out that there exists a lot of interest variance within an environment (Nye, Perlus, & Rounds, 2018; Schelfhout et al., 2019; Tracey et al., 2012), despite the homogeneity that originates through the processes of attraction, selection and attrition (Schneider, 1987; Schneider et al., 2000). Furthermore, literature was in need of methodology that takes this variance into account when generating RIASEC environment profiles (Nauta, 2010; Nye, Perlus, & Rounds, 2018). The present study therefore explored the possibility of using polynomial regression for RIASEC environment profile generation. RIF6 shows a lot of promise, as the method establishes a set of standardized RIASEC coefficients for a specific study program. In doing so, the RIF6 method delineates the program environment similar to methods that use the students in a study program as incumbents (Allen & Robbins, 2010). To this extent, the RIF6 method regresses outcomes like GPA on RIASEC student data within one program. However, the RIF6 method differs from most incumbent methods by also taking into account the student interest RIASEC variances alongside the more common student interest RIASEC averages, as was requested in literature (Nauta, 2010; Nye, Perlus, & Rounds, 2018). RIF6 also shows no indications of multicollinearity that can cause a threat to the reliability of the estimated coefficients, in contrast to other forms of polynomial regression like RIF12 or RIF30. Reliable coefficients are important in order to interpret the RIASEC values in environment profiles. These findings are crucial for literature on environment RIASEC profile generation as polynomial regression not only considers the variance within the environment, but uses the variance of the students within a study program to the full extent, as the constraints on the data are minimal compared to other methods (Edwards, 1993; Edwards, 1994; Edwards & Parry, 1993). The cost of using RIF6 (vs. RIF12) remained limited to about a 2 % loss of accuracy and a drop of $r = 0.36$ to $r = 0.32$ in the correlation with academic achievement.

4.2. Practical implications

As a first major practical implication of our study, RIF6 extracts specific study program environment profiles based on the RIASEC scores and academic achievement of former students. For a practical example that may be useful to study counselors, we refer to the Appendix. As such, RIF6 allows generating interpretable program profiles that profit from the benefits of polynomial regression (i.e., less constraints). The practical implications of using the set of standardized regression coefficients generated by RIF6 as RIASEC program profile are substantial towards interpretation of the RIASEC dimensions. Indeed, standardized RIASEC dimensions (i.e., on a scale with $M = 0$ and $SD = 1$) allow for uniform interpretation and comparison (i.e., higher positive scores or lower negative scores) of the RIASEC dimensions within and between study programs. As an example, the psychology program (see Table 3) is characterized by a very high social dimension compared to the other dimensions. In contrast, program profiles generated by averaging out student RIASEC scores, do not have such power. Averaged profiles do not take into account the variance present in the student population. Even after standardization (i.e., dividing the average score by the standard deviation), the averaged profiles still cannot establish if a high environment profile score on a specific dimension is meaningful, because the criterion is absent. In other words, a high averaged score on a specific dimension for a specific program does not necessarily mean that the specific dimension is environment-determining. In contrast, RIF6 measures the effects of the dimension variances towards achievement so the resulting effects are observed through (individual) effect sizes like an R^2.

As a second major practical implication of our study, RIF6 allows to calculate the PE interest fit between any (future) student and any specific program (or major). For a practical example, we again refer to the Appendix. The resulting fit estimates can then be used to provide an advice set, showing the most fitting programs for an individual student. Together with the profiles of the possible programs, this combination conveys easy-to-interpret study counseling information, for both the counselor and the individual student. In this way, RIF6 can clearly enhance interest-based study counseling, given that the method reflects the relation between study interests and academic achievement more accurately compared to existing methods. Moreover, RIF6 also improves the RIF30 method by matching the student to a set of programs based on the interest variance present in each program separately and not on the variance of all programs combined. In line with literature, each program environment profile is thus characterized by different patterns of vocational interests in relation to academic achievement (Smart et al., 2000) and each program environment will attract students with similar interest profiles (Holland, 1997; Schneider et al., 2000).

4.3. Limitations and future directions

The present study aimed to introduce and validate a new approach to estimate vocational interest fit in an educational context. For this purpose, we relied on data collected in the context of a large and ongoing study orientation project implemented in a Western European university. Although this study context offers several methodological strengths, such as the availability of a large sample of students that can be tracked prospectively over the course of a complete academic year, this context also has limitations. Most obviously, future research is needed that evaluates the generalizability of our findings by comparing our results to those obtained in different educational settings where (1) study programs are composed in a different manner, (2) a different measure is used to assess (future) students' interest profiles and/or (3) student performance (i.e., first year GPA) is assessed differently, for instance through (future) degree attainment. Studies towards degree attainment seem especially warranted, as a degree forms the primary gateway to the work field (Schelfhout, Wille, Fonteyne, Roels, Derous, et al., 2021).

The present study also assessed the strength of the association between interest fit and academic performance relative to the predictive power of self-reported (global) performance in high school. Although there is evidence supporting the validity of such self-reports (Schelfhout et al., 2022), future research can also consider a broader and more differentiated set of alternative predictors of academic achievement when evaluating the performance of the presented RIF methods.

In addition to these study limitations, the interpretation of our results is also limited by a number of characteristics of the featuring RIF. First, RIF calculates PE interest fit (i.e., for individuals) and creates profiles (i.e., for environments) in function of a specific criterion. We specifically relied on GPA as a criterion in the present study given its relevance from a theoretical, methodological and practical perspective. Theoretically, the present study addressed the fundamental question to what extent interest fit indeed predicts successful performance in environments, as predicted by PE fit theory (Holland, 1997; Schneider, 1987; Schneider et al., 2000). Methodologically, the GPA dependent variable was suited nicely to operationalize the academic achievement criterion as (curvi-) linear polynomial regression demands a continuous dependent variable (Edwards, 1993; Edwards & Parry, 1993; Edwards, 1994; Van Iddekinge et al., 2011). And practically, an important aim of study counseling is to heighten the chances that students are oriented towards study programs they can complete successfully (Schelfhout et al., 2022). First-year GPA is arguably accepted as one of the most predictive variables towards degree attainment (Schelfhout et al., 2022; Schneider & Preckel, 2017). Yet, other criteria of PE fit besides pure performance (i.e., first year GPA) can also prove valid, like retention (Van Iddekinge et al., 2011), satisfaction (Hoff et al., 2020; Wiegand et al., 2021) and wellbeing (van Vianen, 2018).

Second, the results from the present study need to be interpreted while considering the nature of the study programs that were included. Specifically, the educational system featuring in the present study allowed for a straightforward delineation of 22 study programs, because each of these programs has a different focus, while the curricula within these programs was highly similar for all students enrolled. Other educational systems (or universities within these systems) potentially offer a less clearly delineated set of programs from which students can choose; future research can evaluate the validity and psychometric properties of the proposed methodology in such contexts. For really small populations (i.e., not enough students to regress twelve or even six RIASEC terms on an outcome) researchers can consider pooling similar environments as an R^2 will always give an indication of how well the vocational interests fit the outcome in the pooled environment. For example, the psychology program from the present study is relatively heterogeneous as the program comprises many related, but also different courses (which everyone has to take), while still being considered one program. Future research can help to systematically evaluate the consequences of such environment heterogeneity on the performance of the RIF methodology. Although the proposed RIF methods yield robust estimated program parameters when applied to the large dataset in the present study, we acknowledge that these RIF estimates are sample-dependent, which is an inherent challenge of polynomial regression (see Xu & Li, 2020). For the present study, the question thus remains to which extent the RIF estimates can generalize to similar programs at different higher education institutions (i.e., colleges or universities). However, the purpose of the current study was primarily to investigate the characteristics of program-specific RIF, such that the RIF method becomes more accessible to institutions interested in using these procedures. Rather than generalizing parameters to different contexts, we therefore recommend applying the outlined principles directly to these different contexts, using institution-specific data, in order to enhance study orientation.

Finally, although restricted to the educational context in the present study, RIF can also be applied to other contexts in which interest-based matching is relevant. Specifically, when the combination of interest and performance (or also satisfaction and retention) data is available in work contexts, RIF6 and RIF12 can leverage an optimal match between workers and occupational environments based on interest fit. Further research on the RIF method can explore the conditions for a successful application in the work field.

5. Conclusion

The present study validates RIF6 and RIF12 as additional applications of polynomial regression for calculating PE interest fit in the delineated environment of a specific study program. By addressing the limitations of polynomial regression, RIF6 and RIF12 can detect an even stronger positive effect between PE interest fit and academic achievement for a specific study program. In doing so, RIF6 can also generate reliable study program profiles by taking into account the interest variance of students in these programs. Ultimately, these program profiles can be used to calculate the PE interest fit between students and study programs, in order to guide individual students towards fitting study programs in higher education.

CRediT authorship contribution statement

Stijn Schelfhout: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Mona Bassleer:** Conceptualization, Writing – review & editing. **Bart Wille:** Conceptualization, Writing – review & editing. **Sofie Van Cauwenberghe:** Conceptualization, Writing – review & editing. **Merel Dutry:** Conceptualization, Writing – review & editing. **Lot Fonteyne:** Data curation, Project administration, Software, Writing – review & editing. **Nicolas Dirix:** Conceptualization, Writing – review & editing. **Eva Derous:** Conceptualization, Supervision, Writing – review & editing. **Filip De Fruyt:** Conceptualization, Supervision, Writing – review & editing. **Wouter Duyck:** Conceptualization, Project administration, Supervision, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jvb.2022.103748.

References

Allen, J., & Robbins, S. (2010). Effects of interest–study program congruence, motivation, and academic performance on timely degree attainment. *Journal of Counseling Psychology, 57*(1), 23–35. https://doi.org/10.1037/a0017267

Barbeau, E. J. (2003). *Polynomials*. Springer.

Brown, S. D., & Gore, P. A., Jr. (1994). An evaluation of interest congruence indices: Distribution characteristics and measurement properties. *Journal of Vocational Behavior, 45*, 310–327. https://doi.org/10.1006/jvbe.1994.1038

Camp, C. C., & Chartrand, J. M. (1992). A comparison and evaluation of interest congruence indices. *Journal of Vocational Behavior, 41*, 162–182. https://doi.org/10.1016/0001-8791(92)90018-U

De Cooman, R., de Gieter, S., Pepermans, R., Hermans, S., du Bois, C., Caers, R., & Jegers, M. (2009). Person–organization fit: Testing socialization and attraction–selection–attrition hypotheses. *Journal of Vocational Behavior, 74*(1), 102–107. https://doi.org/10.1016/j.jvb.2008.10.010

Edwards, J. R. (1993). Problems with the use of profile similarity indices in the study of congruence in organizational research. *Personnel Psychology, 46*(3), 641–665. https://doi.org/10.1111/j.1744-6570.1993.tb00889.x

Edwards, J. R. (1994). The study of congruence in organizational behavior research: Critique and a proposed alternative. *Organizational Behavior and Human Decision Processes, 58*, 51–100. https://doi.org/10.1006/obhd.1994.1029

Edwards, J. R., & Parry, M. E. (1993). On the use of polynomial regression equations as an alternative to difference scores in organizational research. *Academy of Management Journal, 36*(6), 1577–1613. https://doi.org/10.5465/256822

Everitt, B. S., & Skrondal, A. (2010). *The Cambridge dictionary of statistics*. Cambridge University Press.

Fonteyne, L., Wille, B., Duyck, W., & De Fruyt, F. (2017). Exploring vocational and academic fields of study: Development and validation of the flemish SIMON interest inventory (SIMON-I). *International Journal for Educational and Vocational Guidance, 17*(2), 233–262. https://doi.org/10.1007/s10775-016-9327-9

Guan, Y., Deng, H., Fan, L., & Zhou, X. (2021). Theorizing person-environment fit in a changing career world: Interdisciplinary integration and future directions. *Journal of Vocational Behavior, 126*. https://doi.org/10.1016/j.jvb.2021.103557

Hoff, K. A., Chu, C., Einarsdóttir, S., Briley, D. A., Hanna, A., & Rounds, J. (2021). Adolescent vocational interests predict early career success: Two 12-year longitudinal studies. *Applied Psychology, 71*, 49–75. https://doi.org/10.1111/apps.12311

Hoff, K. A., Song, Q. C., Wee, C. J. M., Phan, W. M. J., & Rounds, J. (2020). Interest fit and job satisfaction: A systematic review and meta-analysis. *Journal of Vocational Behavior, 123*, Article 103503. https://doi.org/10.1016/j.jvb.2020.103503

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

Jackson, S. (2017). *Statistics plain and simple* (4th ed.). Cengage Learning.

Johnston, R., Jones, K., & Manley, D. (2018). Confounding and collinearity in regression analysis: A cautionary tale and an alternative procedure, illustrated by studies of british voting behaviour. *Quality & Quantity, 52*(4), 1957–1976. https://doi.org/10.1007/s11135-017-0584-6

Kennedy, P. (2003). *A guide to econometrics*. MIT Press.

Kraha, A., Turner, H., Nimon, K., Zientek, L. R., & Henson, R. K. (2012). Tools to support interpreting multiple regression in the face of multicollinearity. *Frontiers in Psychology, 3*, 44. https://doi.org/10.3389/fpsyg.2012.00044

Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior, 45*(1), 79–122. https://doi.org/10.1006/jvbe.1994.1027

Maxwell, S. E., Kelley, K., & Rausch, J. R. (2008). Sample size planning for statistical power and accuracy in parameter estimation. *Annual Review of Psychology, 59*, 537–563. https://doi.org/10.1146/annurev.psych.59.103006.093735

Nauta, M. M. (2010). The development, evolution, and status of Holland's theory of vocational personalities: Reflections and future directions for counseling psychology. *Journal of Counseling Psychology, 57*(1), 11–22. https://doi.org/10.1037/a0018213

Nye, C. D., Butt, S. M., Bradburn, J., & Prasad, J. (2018). Interests as predictors of performance: An omitted and underappreciated variable. *Journal of Vocational Behavior, 108*, 178–189. https://doi.org/10.1016/j.jvb.2018.08.003

Nye, C. D., Perlus, J. G., & Rounds, J. (2018). Do ornithologists flock together? Examining the homogeneity of interests in occupations. *Journal of Vocational Behavior, 107*, 195–208. https://doi.org/10.1016/j.jvb.2018.04.004

Nye, C. D., Prasad, J., Bradburn, J., & Elizondo, F. (2018). Improving the operationalization of interest congruence using polynomial regression. *Journal of Vocational Behavior, 104*, 154–169. https://doi.org/10.1016/j.jvb.2017.10.012

Nye, C. D., Su, R., Rounds, J., & Drasgow, F. (2012). Vocational interests and performance. A quantitative summary of over 60 years of research. *Perspectives on Psychological Science, 7*(4), 384–403. https://doi.org/10.1177/1745691612449021

OECD. (2017). The state of higher education 2015–2016. Retrieved, June 6th 2019 from http://www.oecd.org/education/imhe/the-state-of-higher-education-201516.htm.

Oh, I.-S., Han, J. H., Holtz, B., Kim, Y. J., & Kim, S. (2018). Do birds of a feather flock, fly, and continue to fly together? The differential and cumulative effects of attraction, selection, and attrition on personality-based within-organization homogeneity and between-organization heterogeneity progression over time. *Journal of Organizational Behavior, 39*(10), 1347–1366. https://doi.org/10.1002/job.2304

Rounds, J., & Tracey, T. J. G. (1996). Cross-cultural structural equivalence of RIASEC models and measures. *Journal of Counseling Psychology, 43*(3), 310–329. https://doi.org/10.1037/0022-0167.43.3.310

Rounds, J., Smith, T., Hubert, L., Lewis, P., & Rivkin, D. (1999). Development of Occupational Interest Profiles (OIPs) for the O*NET part 1: Report. Raleigh, NC: National Center for O*NET Development. https://www.onetcenter.org/reports/OIP.html.

Sackett, P. R., Laczo, R. M., & Lippe, Z. P. (2003). Differential prediction and the use of multiple predictors: The omitted variables problem. *Journal of Applied Psychology, 88*(6), 1046–1056. https://doi.org/10.1037/0021-9010.88.6.1046

Sawilowsky, S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods, 8*(2), 467–474. https://doi.org/10.22237/jmasm/1257035100

Schelfhout, S., Wille, B., Fonteyne, L., Roels, E., De Fruyt, F., & Duyck, W. (2019). The effects of vocational interest on study results: Student person – environment fit and program interest diversity. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0214618

Schelfhout, S., Wille, B., Fonteyne, L., Roels, E., De Fruyt, F., & Duyck, W. (2021a). From interest assessment to study orientation: An empirical advice set engine. *Journal of Experimental Education*. https://doi.org/10.1080/00220973.2019.1605327

Schelfhout, S., Wille, B., Fonteyne, L., Roels, E., Derous, E., de Fruyt, F., & Duyck, W. (2021b). How interest fit relates to STEM study choice: Female students fit their choices better. *Journal of Vocational Behavior, 129*. https://doi.org/10.1016/j.jvb.2021.103614

Schelfhout, S., Wille, B., Fonteyne, L., Roels, E., Derous, E., de Fruyt, F., & Duyck, W. (2022). How accurately do program-specific basic skills predict study success in open access higher education? *International Journal of Educational Research, 111*. https://doi.org/10.1016/j.ijer.2021.101907

Schneider, B. (1987). The people make the place. *Personnel Psychology, 40*(3). https://doi.org/10.1111/j.1744-6570.1987.tb00609.x

Schneider, B., Smith, D. B., & Goldstein, H. W. (2000). Attraction–selection–attrition: Toward a person–environment psychology of organizations. In W. B. Walsh, K. H. Craik, & R. H. Price (Eds.), *Person–environment psychology: New directions and perspectives* (pp. 61–85). Lawrence Erlbaum Associates Publishers.

Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological Bulletin, 143*(6), 565–600. https://doi.org/10.1037/bul0000098

Sheather, S. (2009). *A modern approach to regression with R*. New York, NY: Springer.

Smart, J. C., Feldman, K. A., & Ethington, C. A. (2000). *Academic disciplines: Holland's theory and the study of college students and faculty*. Vanderbilt University Press.

Su, R., & Rounds, J. (2015). All STEM fields are not created equal: People and things interests explain gender disparities across STEM fields. *Frontiers in Psychology, 6*, 189. https://doi:10.3389/fpsyg.2015.00189.

Su, R., Zhang, Q., Liu, Y., & Tay, L. (2019). Modeling congruence in organizational research with latent moderated structural equations. *Journal of Applied Psychology, 104*(11), 1404–1433. https://doi.org/10.1037/apl0000411

The Bologna Declaration. (1999). Joint declaration of the European Ministers of Education. June 19, 1999. Retrieved from https://www.eurashe.eu/library/.

Tinsley, H. E. A. (2000). The congruence myth: An analysis of the efficacy of the person-environment fit model. *Journal of Vocational Behavior, 56*(2), 147–179. https://doi.org/10.1006/jvbe.1999.1727

Tracey, T. J. G., & Rounds, J. (1996). The spherical representation of vocational interests. *Journal of Vocational Behavior, 48*(1), 3–41. https://doi.org/10.1006/jvbe.1996.0002

Tracey, T. J. G., Allen, J., & Robbins, S. B. (2012). Moderation of the relation between person-environment congruence and academic success: Environmental constraint, personal flexibility and method. *Journal of Vocational Behavior, 80*(1), 38–49. https://doi.org/10.1016/j.jvb.2011.03.005

Van Iddekinge, C. H., Roth, P. L., Putka, D. J., & Lanivich, S. E. (2011). Are you interested? A meta-analysis of relations between vocational interests and employee performance and turnover. *Journal of Applied Psychology, 96*(6), 1167–1194. https://doi.org/10.1037/a0024343

van Vianen, A. E. M. (2018). Person-environment fit: A review of its basic tenets. *Annual Review of Organizational Psychology and Organizational Behavior, 5*, 75–101. https://doi.org/10.1146/annurev-orgpsych-032117-104702

Wiegand, J. P., Drasgow, F., & Rounds, J. (2021). Misfit matters: A re-examination of interest fit and job satisfaction. *Journal of Vocational Behavior, 125*. https://doi.org/10.1016/j.jvb.2020.103524

Wille, B., Tracey, T. J. G., Feys, M., & De Fruyt, F. (2014). A longitudinal and multi-method examination of interest-occupation congruence within and across time. *Journal of Vocational Behavior, 84*(1), 59–73. https://doi.org/10.1016/j.jvb.2013.12.001

Xu, H., & Li, H. (2020). Operationalize interest congruence: A comparative examination of four approaches. *Journal of Career Assessment, 28*(4), 571–588. https://doi.org/10.1177/1069072720909825

Young, G., Tokar, D. M., & Subich, L. M. (1998). Congruence revisited: Do 11 indices differentially predict job satisfaction and is the relation moderated by person and situation variables? *Journal of Vocational Behavior, 52*(2), 208–223. https://doi.org/10.1006/jvbe.1997.1587