Predicting Vocational Interests through an AI-based Chatbot

by

Jingyi Li

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Auburn, Alabama
August 5, 2023

Keywords: artificial intelligence, chatbot,
machine learning, vocational interests, psychometric properties

Approved by

Jinyan Fan, Chair, Professor, Psychology
Alejandro Lazarte, Associate Professor, Psychology
Jesse Michel, Associate Professor, Psychology

Abstract

Emerging research has provided compelling evidence that vocational interests serve as robust predictors of various life and work-related outcomes. However, vocational interests assessment did not see much innovation since its inception and still relies heavily on the traditional self-report method. The current study investigated the feasibility of assessing vocational interests using an AI chatbot, where user-generated text inputs were gathered and analyzed utilizing machine learning algorithms. I examined the psychometric properties of machine-inferred interests scores, namely their (1) internal consistency, (2) split-half reliability, (3) factorial validity, (4) convergent validity and discriminant validity, and (5) criterion-related and incremental validity. Participants in the training sample (n = 229) were recruited from the SONA system at a large public university in Florida, while participants in the independent test sample (n = 88) were recruited from the Business School of a large public Southeastern university. Participants first interacted with an AI chatbot for approximately 45 minutes and then completed an online vocational interests measure on Qualtrics. They were then asked to report their college GPA, ACT or/and SAT scores, and high school GPA, and completed surveys that measured major satisfaction, withdrawal intentions, absenteeism, health complaints, and subjective physical health. The results provided encouraging evidence for machine-inferred vocational interests scores in terms of internal consistency, split-half reliability, factorial validity, and convergent and discriminant validity, but failed to support the criterion-related and incremental validity.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AI	Artificial Intelligence
MA	Machine-Inferred
ML	Machine Learning
RMSE	Root Mean Square Error
SR	Self-Reported

Predicting Vocational Interests through an AI-based Chatbot

Machine learning (ML) has been a hot topic recently and has been applied in various areas of psychological research. Increasing attention has been focused on using machine learning to predict the human mind, especially personality (e.g., Hickman et al., 2021; Youyou et al., 2015). Those research results showed a promising future for machine learning in measuring individual differences. Meta-analyses conducted to estimate correlations between machine-inferred personality scores derived from digital trace data and traditional Big Five self-reported personality scores indicate that correlations range from .29 to .40 across the Big Five personality traits, which may be a reasonable level of convergence considering the noisiness of trace data (Azucar et al., 2018; Settanni et al., 2018). However, even though ML has the potential to predict various types of individual differences, personality received most of the attention from psychology researchers while other individual differences are relatively under-studied. Vocational interests, which reflect relatively stable individual differences in motives, goals, and personal strivings both within and beyond the vocational domain (Stoll et al., 2017; Van Iddekinge et al., 2011), hold a crucial position among various types of individual differences in applied psychology (Chernyshenko et al., 2010). I argue that vocational interests are important noncognitive factors that influence major work and life outcomes, and their assessment deserves innovation. In this research, I explore the application of machine learning in predicting vocational interests. Specifically, I build ML models using chatbot text scripts to predict vocational interests and examine the following psychometric properties of machine-inferred career interests scores: internal consistency, split-half reliability, factorial validity, convergent and discriminant validity, and criterion-related and incremental validity.

Theories of Vocational Interests

Holland's (1997) theory of vocational personalities and environments is widely regarded as the predominant theory in the field of vocational interests. The core argument of the theory is that both individuals and their environments can be characterized using the RIASEC model, which encompasses six distinct types of vocational interests: *Realistic, Investigative, Artistic, Social, Enterprising,* and *Conventional*. Another important element of the theory is the congruence hypothesis, which assumes that people find the environments satisfying when the environmental patterns are congruent with their own interest patterns. Holland’s (1997) model also specifies a two-dimensional hexagonal structure of interrelations among the six interest types such that interest types adjacent to each other have the strongest correlations, those that alternate to each other are less correlated, and those opposite to each other are less likely to be correlated.

Su et al. (2019) recently developed an eight-dimension interest model, the SETPOINT model (SETPOINT: *H*ealth *S*cience, *C*reative *E*xpression, *T*echnology, *P*eople, *O*rganization, *I*nfluence, *N*ature, and *T*hings), by conducting an extensive analysis of existing basic interest scales. In this model, three levels of interests are structured hierarchically, with general interest dimensions that describe the general characteristics of individuals and environments at the highest level, basic interests that reflect the same classes of activities at the intermediate level, and the specific work activities at the lowest level (Su et al., 2019). In addition, by incorporating most existing basic interest scales, the SETPOINT model comprehensively incorporated not only occupations that have served the creation of RIASEC model over the past century, but also the ones that flourished recently, such as those in healthcare and technology, which were only partially represented in the RIASEC model. Another difference between the SETPOINT and the

RIASEC model is that the former is a dimensional model, which means there are no specified interrelationships between the eight interest dimensions.

Several other theories also describe vocational interests and their link to occupations. For example, Tracey and colleagues developed a spherical model representing the three-dimensional structure of vocational interests (Tracey, 2002; Tracey & Rounds, 1996). Like Holland's (1997) model, the spherical model also highlights the structured relationships among the eight proposed interest types. Additionally, Gottfredson's (2005) theory of circumscription and compromise underscores the significance of interests as key motivational factors that influence individuals' career choices and pursuits. The theory suggests that people match themselves to careers by four processes: growing in cognitive ability, increasing self-directed development, eliminating least favored vocational interests, and accommodating to external constraints. However, I will focus my attention on Su et al.' (2019) SETPOINT model in this research for two reasons. First, the SETPOINT model is based on a recent extensive analysis of existing basic interest scales, which endows the model with timeliness and comprehensiveness. In addition, due to its hierarchical nature, the SETPOINT model incorporates both broad levels of interest measurement and specific measures of basic interests. The former provides a concise organizing framework that could guide research, while the latter, with lower levels of measurement (i.e., more specific, and more refined), allows for more detailed investigations into the relationship between interests and various outcomes as well as provides stronger predictive power, as shown by research evidence (Hanna & Rounds, 2020; Song et al., 2022; Su et al., 2019; Van Iddekinge et al., 2011).

Vocational interests assessment has long been used to aid educational and career decision-making (Strong, 1943; Zickar & Min, 2019). Both researchers and practitioners have put forth the suggestion that individuals derive greater enjoyment and derive more benefits from their work when there is congruence between their personal characteristics and the requirements of the occupation (Barrick & Mount, 1991; Holland, 1966; Super, 1953). Nowadays, students have access to interests assessments in high school and college counseling services. In addition, various settings such as employment agencies, social services and community agencies, private organizations, and online resources such as O*NET all provide interests assessments (Hansen, 2019). Commensurate with its extensive use for guidance purposes in various settings, vocational interests are indeed the strongest known predictors of career choices with an estimated overall hit rate of 50.8% for predicting career choice (Hanna & Rounds, 2020).

The power of interests is not limited to career guidance. Recent research has found that vocational interests can predict persistence and performance in both work and academic contexts (Nye et al., 2012; Nye et al., 2017; Van Iddekinge et al., 2011), and work satisfaction (Assouline & Meir, 1987; Tsabari et al., 2005; Nye et al., 2018). Though some research did not find interest as a strong predictor of satisfaction (Assouline & Meir, 1987; Tsabari et al., 2005), meta-analysis shows that the correlations for performance outcomes can reach up to 0.32 when interest congruence was used as the predictor (Nye et al., 2017). Indeed, interest assessment is employed within the U.S. military for recruitment and placement objectives, as highlighted by Ingerick and Rumsey (2014). Moreover, researchers have been advocating for the integration of interests assessment into the selection process to predict various factors such as occupational membership, job performance, and turnover (Kirkendall et al., 2020; Nye et al., 2012; Nye et al., 2019; Su et al., 2019; Van Iddekinge et al.,2011).

With all the benefits in real-world settings, recent progress in the interests literature also indicates the theoretical value of vocational interests in expanding the scope of the predictive ability of individual differences. Stoll et al. (2017) found that vocational interests were significant predictors of three work-related outcomes (i.e., full-time employment, gross income, and unemployment), three relationship outcomes (i.e., getting married, having children, and never being in a relationship), and one health-related outcome (i.e., perceived health status). In the study conducted by Stoll et al. (2017), it was also found that vocational interests were more potent predictors of work-related and relationship outcomes compared to the Big Five personality traits. Stoll et al.'s study updated our knowledge of vocational interests as important noncognitive factors that need more research attention.

Vocational interests are usually measured using self-report measures. Some notable examples of interest measures include Vocational Preference Inventory (VPI; Holland, 1985), Self-Directed Search (SDS; Holland et al., 1994), O*NET® Interest Profiler (Rounds et al., 2021), Kuder Preference Record (Kuder, 1983), and Strong Interest Inventory (Donnay et al., 2005). Despite being an important concept in both applied and academic settings, there are barely any innovations in interest measures since the first one has been developed nearly a century ago. According to Chernyshenko et al. (2019), the majority of current interest inventories rely on text-based stimuli, such as checklists of occupations or activity descriptors. Self-report vocational interests measures can be long and tedious and are also vulnerable to faking when used in applicant samples. I argue that the machine learning approach will provide various advantages over the traditional self-report interest assessment, such as being efficient in data management, less dependent on self-rating, and less prone to faking. In the subsequent section, I will present an overview of how the machine learning (ML) approach can be applied to the measurement of vocational interests. Additionally, I will delve into a more detailed exploration of the advantages associated with using ML in this context.

The ML Approach

Machine learning is a type of artificial intelligence (AI), which focuses on using algorithms to learn from data to make decisions and predict outcomes with minimal human intervention (IBM Cloud Education, 2020). To predict individual differences such as personality, values, or interests using the ML approach, two steps are usually involved. The first step is to build predictive models in a large sample, in which data will be collected both on the predictor side and the criteria side. Predictors in ML models are features extracted from digital footprints such as social media posts (e.g., Kern et al., 2019), Facebook likes (e.g., Kosinski et al., 2013; Youyou et al., 2015), facial expressions (e.g., Hickman et al., 2021), etc. These features are potentially relevant to the outcomes, and they are identified automatically or manually from the digital footprints. The criteria used to build the predictive models are usually objective outcomes (e.g., career choice, Kern et al., 2019) or self-reported questionnaire scores (e.g., self-reported personality scores, Kosinski et al., 2013; Youyou et al., 2015). The establishment of empirical links between the features (indicative of vocational interests) and the criteria (outcomes or variables of interest) can be accomplished through various analytical methods. These methods may include linear regression analysis, logistic regression analysis, neural networks, or other suitable statistical techniques (Fan et al., 2023). K-fold cross-validation is usually used to avoid model overfitting (Hickman et al., 2021). After the model parameters are obtained in this sample, a cross-sample validation may be conducted using an independent sample. The purpose of this step is to cross-validate the predictive models using a new independent sample to generate more rigorous models. The second step is to apply the trained model to new individuals without self-reported questionnaire scores to predict their scores on the construct (e.g., personality, vocational interests, etc.)

The most explicit advantage of the ML approach is its efficiency. On one hand, once the machine learning (ML) models have been constructed, it becomes possible to infer various individual differences automatically and simultaneously (Fan et al., 2023), such as personality traits, values, vocational interests, competencies, and more, using a single set of data inputs. Such property enables efficient examination of a single individual difference or a combination of individual differences in one study. For instance, by using ML models, Kern et al. (2019) inferred Twitter users' psychological profiles (i.e., personality and values) automatically and simultaneously via their Twitter posts. On the other hand, ML could manage data from a large sample as well as periodically receive new data and update the model parameters (Gonzalez et al., 2022). When being used to predict vocational interests, such a feature enables the ML approach to learn from experience and adjust to the rapidly changing labor market (Gati & Kulcsár, 2021). In summary, ML allows for the efficient management of complex datasets that are difficult to handle otherwise.

Another advantage of the ML approach is that there is no self-rating needed once the models are built. As an issue with many traditional measures, self-reported measures are sometimes long and tedious to complete, which could lead to less diligent responses (Tracey, 2020). In vocational interest measures, basic interest measures, which usually have higher predictive validity than broad interest measures, are usually much longer due to more specific interest categories being included compared to the latter (Song et al., 2022). However, the ML approach could improve the assessment experience of test takers as it does not rely on test takers to reflect and evaluate themselves on hundreds of items. Instead, it can capture and analyze data such as digital footprints and dialogues of individuals, which could save or reduce test takers' cognitive load.

The ML approach also has the potential to ameliorate faking issues that are common in self-report measures. On one hand, despite that faking issues were less mentioned in vocational interests assessment literature, given its infrequent use in high-stakes settings, vocational interests have great potential in personnel selection as an indicator of person-environment fit (Nye et al., 2017). On the other hand, applicants can easily fake when they are asked to rate job activities according to their interests if they have some knowledge about the position they are applying for. Previous military research, as indicated by Hough et al. (2001), discovered that soldiers may manipulate their responses when instructed to do so in interest inventories. Although I do not plan to explicitly examine faking issues in the present study, building the models and examining the basic psychometric properties of the ML approach to predicting vocational interests establish the ground for future exploration. In summary, the ML approach has several advantages compared to the self-reported approach, and research should be conducted to explore how such an approach could be used in the vocational interests field (Tracey, 2020).

Despite the above potential advantages, researchers have limited knowledge of the ML approach to predicting individual differences (Fan et al., 2023). One of the challenges that the ML approach faces is the privacy and legality of digital data use (Oswald, 2020; Tracey, 2020). Indeed, certain machine learning (ML) approaches may involve the collection and analysis of users' social media content. While this can provide valuable insights for inferring individual differences, such as interests or personality traits, it also raises privacy concerns and potential legal issues (Fan et al., 2023). Sharing personal social media content can compromise users' privacy, as it involves accessing their online activities and potentially sensitive information. Stoughton et al. (2015) found participants perceived their privacy as being invaded when they were told organizations had evaluated their social media to assess their professionalism. In addition, knowing that their social media is being analyzed, applicants may perceive less procedural justice because they have less control over the selection process when they don't have the opportunities to voice their opinion (Gonzalez et al., 2022). Moreover, the automated process may decrease applicants' motivation as it is less able to satisfy three fundamental needs of humans as stated in self-determination theory (SDT; Deci & Ryan, 2012): autonomy, competence, and relatedness. Specifically, according to Gonzalez et al. (2022), collecting social media information may allow fewer opportunities for applicants to control the process due to their inability to affect organizations' evaluation of them after they give organizations access to their social media accounts. Such an approach may also prohibit applicants from fully demonstrating their competence as they may not think social media could reflect their real job-related qualifications. In addition, by automatically collecting applicants' social media information using the algorithm, applicants don't have the opportunity to establish connections with humans in the assessment process. However, there are other ways that researchers and practitioners collect data for the ML approach, without making users perceive their privacy as being invaded or decreasing their motivation. For example, Hickman et al. (2022) and Leutner et al. (2021) used automated video interviews (AVIs) as the data source for the ML algorithms. Fan et al. (2023) utilized an AI-based chatbot that enabled two-way human like conversations between participants and the chatbot. This study will adopt the same chatbot as an assessment platform to collect data. In the next section, I introduce the AI-based chatbot and summarize unique advantages over other commonly used data sources in ML.

An Introduction to the Juji Chatbot

In the present study, I use an AI-based chatbot provided by Juji (https://juji.io) with which users have one-on-one and two-way communications. I chose Juji's chatbot as our focal data source for several reasons. First, the AI-based chatbot may elicit better applicant reactions compared to the social media approach to assessing individual differences. By imitating humans to conduct conversations, the AI-based chatbot provides access to more control over the assessment process, in which users could engage in two-way interaction with the chatbot. In addition, it provides more opportunities for users to establish connections with the human-like chatbot. By using users' chat scripts throughout the whole process as data input, users also have more opportunities to demonstrate their competence in their free-text responses. Such properties of chatbots may ameliorate the negative applicant reactions I mentioned earlier toward the ML approaches in general. Indeed, research evidence supported Juji's capability to engage in effective interactional conversations and understand human needs and wants (Gupta et al., 2019; Lee et al., 2019), which enables more human-like interactions with users. Second, research has suggested that, in comparison to online surveys, chatbots generate a higher level of participant engagement and yield better-quality responses. Participants are less prone to engage in satisficing behavior when they are interacting with chatbots (Kim et al., 2019). Research found that Juji's chatbot interactions tend to produce responses that are more informative, relevant, specific, and clear (Xiao et al., 2020). Third, Juji's AI-based chatbot allows the easy design of a chatbot tailored to the researchers' needs. In the present study, I use Juji's AI chatbot which could be built with an intuitive graphical user interface without requiring users to write computer code. The specific questions that the chatbot asks could be designed by the researchers.

Previous Research on ML and Vocational Interests

I conducted a literature search that limited articles to those related to vocational interests and ML. I was only able to find one study that directly uses the ML approach to the prediction of vocational interests. Kiselev et al. (2020) developed ML models that well predicted social network users' vocational interests via the information on social network users' followed group. Specifically, they recruited 1252 participants from the adware of computerized career guidance on the internet and collected information on their group subscriptions, through the official API of the social network, VKontakte, which is Russia's main social network. Participants also completed a 30-item vocational interest questionnaire that represents a 10-interests type model that extends Holland's RIASEC model (Kiselev et al., 2020). Categorical features that were extracted from those group subscriptions were used as predictors of the self-reported vocational interests types (i.e., the criteria). Five-fold cross-validation was used to avoid model overfitting, but no independent test sample was recruited. The prediction accuracy was indexed by the match of each participant's self-reported first-rank vocational interest and the machine-inferred first-rank vocational interest. The prediction accuracies range from 0.68 to 0.85 for each of the 10 interest types, which indicates reasonably good prediction performance (Kiselev et al., 2020). This research demonstrates a good exploration of the feasibility of ML in predicting vocational interests. However, it is important to acknowledge that the study may have limited generalizability due to the specific context and sample used. The models in the study were constructed based on a Russian vocational interest model and utilized social network information from a Russian sample. As a result, the findings may not be directly applicable or generalizable to other cultural or geographical contexts. In addition, although five-fold cross-validation was used to avoid model overfitting, no independent cross-sample validation was conducted to further support the prediction performance of the models. Moreover, the reliability and other aspects of validity of machine-inferred interests scores were not reported.

Another line of research, though not directly related to our topic, provides valuable information on how ML could be used to predict career choices and give career guidance. A representative work is Song et al. (2022), which developed an ML-augmented method to aid the match of people and jobs using data from a TIME Magazine online survey. Traditionally, the matching between people and occupation is achieved in several ways, such as comparing the high-point code of the person's interest profile and the occupation's interest profile and there is a match if the two codes are the same (e.g., I in IRASEC and I in IASERC), calculating the Euclidean distance between the two profiles, or using the profile correlation. In other words, the traditional approach predicts people's career choices or gives career guidance based on the theoretical match of interest types between people and occupations. However, the ML-augmented method builds models that use participants' interest scale scores and item scores to predict the interest scores of their occupations such that the predictors are participants' interest item scores and scale scores, and the criteria are the interest scores of their occupations. In other words, the ML-augmented method predicts people's occupations based on empirical links between people and occupations found in a specific sample. Their study is different from my approach in that, in their study, ML is used after vocational interests scores are obtained to increase the accuracy of interest inventories in predicting career choices.

Similarly, Kern et al. (2019) used ML in predicting Twitter users' occupations based on their Twitter posts. Specifically, they used ML in two steps. They first obtained the latest 200 tweets of 1.5 million Twitter users and inferred the normalized personality trait and value scores (i.e., psychological profile) using IBM Watson's system. Next, they trained an ML algorithm to learn the relationship between users' psychological profiles and their job titles as indicated in their Twitter profiles. The accuracy of this model in predicting users' occupations was tested in a second set of data. The accuracies obtained were all higher than 70% for each of their classifiers (i.e., using only 5 traits, using only 5 values, or using all 10 features), which demonstrates good performance, too. However, in this line of research, ML was used to predict career choices and psychological profiles, not interests scores themselves. I argue that examining the use of the ML approach to predicting interests scores brings more theoretical and practical benefits than predicting career choices. On the theoretical side, such a query allows us to explore the feasibility of a new measurement method for vocational interests by examining its psychometric properties. On the practical side, machine-inferred interests scores, once proven its validity, could be used to better predict various life and work-related outcomes in addition to career choices, such as performance, satisfaction, health status, etc.

In summary, research on using the ML approach to predicting vocational interests is strikingly scarce. I was unable to identify any article that directly taps into this issue in high-quality psychology journals. Existing research either does not comprehensively examine the validity of the ML approach, lacks generalizability due to data and sample limitation, does not directly use the ML approach in predicting vocational interests scores, or lacks emphasis on or scientific description of how the ML models were built or validated. Those issues together highlight the need for research that (1) directly uses the ML approach to predicting vocational interests scores, (2) describes the ML model building and testing steps in detail in scientific languages, and (3) examines the psychometric properties of the ML models comprehensively.

Examining the Validity of Machine-Inferred Vocational Interests Scores

The present study builds ML models that predict vocational interests scores based on participants' text input in an AI-based chatbot. Furthermore, consistent with the recommendations of Bleidorn and Hopwood (2019) regarding the assessment of construct validity in machine learning (ML) approach for assessing individual differences, the current study investigates the psychometric properties of ML models across various dimensions. These dimensions include internal consistency, split-half reliability, factorial validity, convergent validity, discriminant validity, as well as criterion-related and incremental validity. To my best knowledge, the proposed research is the first study developing and systematically assessing the machine learning approach in predicting vocational interests, by developing a cognitive, AI-based chatbot approach, and validating it through comprehensive examinations of its psychometric properties. The present study also has significant implications for practice. Given that interest assessments play a crucial role in career guidance and development for a large population, even a small increase in validity achieved through the utilization of a machine learning (ML) approach holds significant importance (Song et al., 2022). The impact of such an increase in validity should not be underestimated, as it has the potential to influence career-related decisions and choices for millions of individuals.

Method

Sample and Procedure

A total of 377 participants were recruited. Two hundred and fifty participants were recruited as the training sample in this study, all of whom were undergraduate students enrolled in different psychology courses. The recruitment was conducted through the SONA system, which is a subject pool managed by the Psychology Department at a public university in Florida.

For the test sample, one hundred and twenty-seven undergraduate students were recruited through several courses in the Business School in a Southeastern public university. The training sample was used to build the model to predict participants' vocational interest scores based on the text input in the virtual chats. The test sample was used to cross-validate the predictive model.

All data collection was completed online. During the study, participants were presented with a series of questions related to various interests topics. These questions, listed in Appendix A, were displayed on a computer screen sequentially, and participants were instructed to enter their responses into a chat box provided on the screen. Next, participants completed an online, self-report vocational interest measure on Qualtrics. Participants' demographic information was collected in this questionnaire, including college major, gender, age, and ethnicity. Each participant was assigned a participant ID which helped match their chat content and their vocational interest scores and criteria data. After that, participants were asked to report their college GPA, ACT or/and SAT scores, high school GPA, and completed five surveys that measured several criterion variables: major satisfaction, withdrawal intentions, absenteeism, health complaints, and subjective physical health. Participants in the training sample were given 3 credit hours after completing the tasks, while participants in the test sample were offered this study as a research opportunity in which they could experience and learn about conducting research.

In the training sample, twenty-one participants who either mistyped their participant IDs, withdrew from the study, had incomplete data (number of chat text words < 430), or failed two attention check questions (Meade & Craig, 2012)) were excluded from the study. I used number of chat text words of 430 as an exclusion criterion of participants' chat text, which is slightly lenient than what is suggested in Juji's built in function (number of chat text words < 500), considering that the sample sizes in this study are small. The remaining 229 participants in the training sample has a mean age of 20.26 and 58% of total are female. The distribution of races is as follows: White (47%), Hispanic (24%), African American (7%), Asian-Eastern (7%), Asian-India (4%), Mixed-race (8%), and other (1%). The top 10 majors with most participants are: Psychology (*n* = 32; 14%), Computer Science (*n* = 22; 10%), Health Science (*n* = 17; 7%), Biology (*n* = 14; 6%), Information Technology (*n* = 13; 6%), Nursing (*n* = 13; 6%), Kinesiology (*n* = 10; 4%), Aerospace Engineering (*n* = 9; 4%), Biomedical (*n* = 7; 3%), Political Science (*n* = 5; 2%).

Thirty-nine participants in the test sample were excluded for the same reasons. The mean age is 21.55 in the remaining 88 participants and 68% of them are female. The distribution of races in the training sample is as follows: White (86%), others are African American (5%), Hispanic (2%), Asian-Eastern (2%), Asian-India (1%), and Mixed-race (3%). The top 10 majors with most participants in them are: Marketing (*n* = 44; 50%), Business (*n* = 20; 23%), Biology (*n* = 13; 15%), Multidisciplinary Studies (*n* = 3; 3%), Psychology (*n* = 2; 2%), Sports and recreation administration (*n* = 2; 2%), Public policy & leadership (*n* = 2; 2%), Finance (*n* = 1; 1%), Political science (*n* = 1; 1%), Journalism (*n* = 1; 1%).

Development of AI Chatbot Interview Questions

I developed a set of interview questions that will be asked by the AI chatbot. Those questions tap into the major components of the definition of vocational interests, with some in open-ended formats while others in close-ended formats. To design those questions, I first referred to the definitions of vocational interests. According to Rounds and Su (2014, p. 98), vocational interests are “trait-like preferences to engage in activities, contexts in which activities occur, and outcomes associated with preferred activities, that motivate goal-oriented behaviors and orient individuals toward certain environments”. Holland (1997) proposed that interest types are preferences for certain activities, skills, abilities, attitudes, and characteristics. People with the same interest type are characterized by similar preferences for tasks, and activities, show abilities in the same areas, and are likely to have similar values and life goals. These definitions highlight several major components related to vocational interests: activities, contexts of the activities, outcomes of the activities, tasks, abilities, values, and life goals. Next, I reviewed the default interview questions provided by Juji and kept those categorized under these major components (e.g., “What do you love to do outside of studying for your classes?”). I next drafted more general questions for those underrepresented components (e.g., values; “Do you think that a job/working is the most important thing in your life?”). While I had developed this set of general questions asking about the participant’s life, I then generated several more specific questions based on theoretic models. Specifically, those questions ask participants to make pairwise comparisons of the eight dimensions of the SETPOINT model and four general value dimensions of Schwartz’s value model (Schwartz, 1992; Schwartz, 2017; e.g., “Could you talk about what is your preference for either of the following types? Conducting research on health-related issues or providing medical services to help cure people OR planning and organizing things/people to facilitate organization activities. You can type "Neither of these interests me" if you like neither of them”). To elicit more responses from participants, follow-up questions asking for more details were added after each major question if there were none from the default Juji questions. All questions will be asked in a predetermined temporal order and are defined by a set of “IF…, THEN…” rules. The interview scripts were finalized and imported to Juji’s no-code AI chatbot builder after three rounds of review by peers and supervisors.

Measures

Self-reported vocational interests. The Comprehensive Assessment of Basic Interests (CABIN; Su et al., 2019) was used as a measurement of the SETPOINT model. CABIN consists of 164 items, with 41 scales and 4 items per scale, representing an eight-dimension model of vocational interest. Each scale represents a basic interest, and each item states an activity related to the basic interest. Participants rated how they feel about engaging in each activity on a 5-point Likert scale ranging from 1 (*dislike a great deal*) to 5 (*like a great deal*).

Machine-inferred vocational interests. The machine-learning algorithm was developed based on the textual features of participants’ chat scripts with the help of the AI firm using the training sample. See the Analysis section for the model-building method.

Vocational interests congruence. Calculating congruence scores needs both individuals’ interest profiles (i.e., scores on each of the SETPOINT dimensions) and the interest profile of the majors that individuals are in. In our study, congruence scores were obtained for individuals with a Marketing or Business major in the test sample, as the number of participants in other majors is not enough for data analysis. The interest profiles of the two majors were the average of all students’ interest profiles in that major. Interests congruence was calculated for each individual by obtaining the correlation between the interest profile of each of them and the interests profile of the major.

Major satisfaction. A 6-item global self-report measure, Academic Major Satisfaction Scale (AMSS; Nauta, 2007) was used to measure global satisfaction with one's major. Each item is rated on a 5-point Likert scale ranging from 1(*strongly disagree*) to 5(*strongly agree*). Nauta (2007) reported strong predictive and convergent validity for the AMSS and a Cronbach's α of .90.

Withdrawal intentions. To measure withdrawal intentions, two items from the scale developed by Schmitt et al. (2008) were utilized. Both items are rated on a 5-point Likert scale from 1(*strongly disagree*) to 5(*strongly agree*). The two items correlated .73 and the alpha was .84 (Schmitt et al., 2008).

Absenteeism. Participants were asked to report the number of times they had missed classes in the past six months for "unavoidable" reasons (i.e., excused absences such as for personal illness) and "unavoidable" reasons (i.e., unexcused absences such as for oversleeping) with five response options ranging from "missed less than 5 times" to "missed more than 30 times" (Schmitt et al., 2008).

Health complaints. A 12-item Physical Symptoms Inventory (PSI) was used, which is a short version of the original 18-item PSI after eliminating 5 items that were rarely endorsed and 1 overlapped item (Spector & Jex, 1998). Each item is a physical, somatic health symptom that could be associated with psychological distress. Participants were asked to rate the frequency they experience each of the symptoms over the past month based on five response choices ranging from "not at all" to "every day".

Subjective physical health. Participants indicated on a 7-point scale from 1 (*much worse*) to 7 (*much better*) how they perceived their current health status compared with other people of the same age and gender (Staudinger et al., 1999).

GPA. Participants reported their accumulative GPA in college.

Major. Participants reported whether they had chosen a major and if so, indicated what major they had chosen.

Control variables. Participants provided their ACT scores, SAT scores, and their high school GPA. During the analysis, high school GPA was employed as a control variable to account for cognitive ability. High school GPA was used in the actual analysis, instead of ACT or SAT scores, due to a larger number of participants reporting high school GPA rather than their ACT or SAT scores. By controlling high school GPA, the study aimed to examine the criterion-related validity of vocational interests in predicting GPA while accounting for the influence of cognitive ability.

Analytic Strategies

Model building. As I discussed earlier, the training sample was used to train the models and the independent test sample was used to cross-validate the trained models. To train the models, natural language processing was used, in which I obtained the text features as predictors for the regression models. First, I segmented the text input of each participant into single sentences, on which the encoder input a tokenized string and outputs a 512-dimensional vector as the sentence embedding (Cer et al., 2018). As a result, for each participant, a set of 512 feature scores was derived, reflecting the averaged sentence embeddings across sentences.

To build the ML models, elastic net regression was used (Zou & Hastie, 2005) on the training sample to learn the relationship between those text features and participants' self-reported vocational interests scores (i.e., CABIN scores). I chose to use elastic net regression because it works well when the number of predictors is large compared to the number of observations (Zou & Hastie, 2005). The elastic net regression models were built in Python 3 using the scikit-learn package (Pedregosa et al., 2011). To tune the models' hyperparameters, I used 5-fold cross-validation, in which the dataset was divided into 5-folds and the prediction models were developed based on 4-folds and then tested on the 5th fold. This process was repeated until each fold had served as the test set once. The final model parameters were obtained by training the models using the complete training sample, incorporating selected hyperparameters that were fixed to the optimal values, as outlined by Fan et al. (2023). After obtaining the trained models, they were then applied to both the training and test samples to predict the basic interests scores. This process generated 41 predictive models in total as I trained one model to predict each basic interest score.

Reliability. Traditionally, it is difficult to examine the internal consistency of machine-inferred scores as the predictors in ML approaches are usually a large number of features that are heterogeneous (Hickman et al., 2021). However, like Fan et al. (2023), I used the basic interest scale scores under each of the eight dimensions of the SETPOINT model as "items" to calculate the internal consistency of dimension scores, as the scale scores have theoretic correlations when they are categorized under the same dimension. To examine the split-half reliability of the machine-inferred scores, participants' text input was randomly split into two halves, and I obtained one set of machine-inferred interests scores for each of the halves. The Spearman-Brown correlations of the two scores were calculated to index the split-half reliability.

Factorial Validity. To assess the factorial validity of the machine-inferred interests scores from the test sample, I employed a statistical approach called Set Exploratory Structural Equation Modeling (Set-ESEM). Set-ESEM, as described by Marsh et al. (2020), is a modeling technique that allows for the examination of two sets of constructs within a single model. In Set-ESEM, researchers can model two sets of constructs simultaneously, and within each set, cross-loadings are permitted among the factors of the same constructs. However, the cross-loadings are constrained to be zero between factors from different sets of constructs.

This property of Set-ESEM enables me to directly estimate the correlation among latent factors derived from machine-inferred and self-reported scores (Fan et al., 2023). To compare the factorial validity of the machine-inferred interests scores and the self-reported interests scores, two aspects of similarity were examined: overall pattern similarity and magnitude similarity. To evaluate the congruence between factors of self-reported interests and machine-inferred interests scores, Tucker's Congruence Coefficients (TCCs) were employed for the former. TCCs provide a measure of pattern similarity between the factor structures of the two sets of scores (Lorenzo-Seva & ten Berge, 2006). Higher TCC values indicate greater congruence or similarity between the factors of self-reported and machine-inferred interests.

On the other hand, the magnitude similarity between the factor loadings of machine-inferred scores and self-reported scores was assessed using Root Mean Square Error (RMSE) calculations. The RMSE quantifies the absolute magnitude of agreement between the factor loadings derived from the machine-inferred scores and those from the self-reported scores. By calculating the RMSE for each interest dimension separately, the researchers can gauge the level of agreement between the factor loadings derived from the machine-inferred scores and the self-reported scores in both the training and testing samples. In addition, two lines that visualize the range of factor loadings differences that are less than .10 were drawn in the plot. The similarity in magnitude were shown by most dots falling within the range (Fan et al., 2023).

Convergent and discriminant validity. I modeled the multitrait-multimethod matrix (MTMM) using Woehr et al.'s (2012) generalizability theory (G-theory) method. To assess the convergent validity and discriminant validity of the machine-inferred interests scores, the methodology described in Woehr et al. (2012) was followed. Several indices were computed, including a convergence index (C1), three discrimination indices (D1, D2, and D2a), as well as two additional correlations.

To provide more specificity, the convergence index (C1) quantifies the expected observed variance resulting from individual differences (person main effects) and the shared variance among trait-method units that is specifically associated with a particular trait. It was computed by averaging the absolute monotrait-heteromethod correlations. A positive and sizable C1 value indicates strong agreement among the measurements, indicating good convergent validity (Woehr et al., 2012).

D1, on the other hand, represents the portion of the anticipated observed variance that is specifically associated with a given trait and is shared among trait-method units. The calculation of D1 involved subtracting the average absolute heterotrait-heteromethod correlations from the monotrait-heteromethod correlations. If D1 value is positive and substantial, the trait-specific variance being assessed is said to outweigh the variance attributed to individual differences (Woehr et al., 2012).

D2 measures the difference between the proportions of anticipated observed variance attributable to trait-specific and method-specific shared variance. The calculation of the discrimination index D2 involved subtracting the average of absolute heterotrait-monomethod correlations from the average of absolute monotrait-heteromethod correlations. A positive and substantial D2 value indicates good discriminant validity, indicating that the variance linked to the traits of interest is more substantial than the variance attributed to the measurement method itself (Woehr et al., 2012).

Considering that machine learning (ML) approaches often have a negative impact on the discriminant validity (Fan et al., 2023; Park et al., 2015; Tay et al., 2020), additional analyses were conducted. To address this concern, the index D2a was calculated by determining the average absolute heterotrait-ML method correlations, which were derived solely from ML methods. Furthermore, to compare the discriminant validity of the ML approach with the traditional self-report method, I compared the average absolute heterotrait-ML method correlations with the average of absolute heterotrait-self-report method correlations. If the latter value surpasses the former, it suggests that the ML approach superior discriminant validity (Fan et al., 2023).

Criterion-related and incremental validity. The study examined the criterion-related validity of machine-inferred interests scores by assessing how well the congruence between these scores and participants' major interests profiles predicted external criteria variables such as major satisfaction, withdrawal intention, absenteeism, health complaints, subjective physical health, and GPA. The study also investigated the incremental validity of interests congruence, determining its unique contribution in predicting outcomes beyond cognitive ability.

To explore the relationship between interest congruence and external criteria, the study initially conducted partial bivariate correlations. These correlations examined the associations between interest congruence and six criterion variables, namely major satisfaction, withdrawal intention, absenteeism, health complaints, subjective physical health, and GPA. Importantly, the analysis controlled for high school GPA to isolate the specific influence of interest congruence on the criterion variables while accounting for cognitive ability. As I mentioned earlier, for each individual, interest congruence is the correlation between the person's interest profile and the interest profile of their major. This analysis only focused on a subsample of students who indicated they have a Business or Marketing major in the survey, as the number of individuals in other majors was not enough for data analysis. Similar to Fan et al. (2023), the study conducted six sets of hierarchical regression analyses. Each analysis used one of the six external criteria (major satisfaction, withdrawal intention, absenteeism, health complaints, subjective physical health, and GPA) as a separate outcome variable. The congruence indices were entered as predictors in the regression models. In those regression analyses, I first regressed the criterion on high school GPA (Step 1), then on congruence indices calculated using self-reported interests scores (Step 2), and then added the congruence indices based on machine-inferred interests scores to the regression model (Step 3). Incremental validity was indexed by significant regression coefficients associated with machine-inferred scores in Step 3.

Results

Reliability

The reliabilities for the 41 basic interest scales and 8 interest dimensions of the self-reported and machine-inferred interest scores are presented in Table 1 and Table 2, respectively. The findings presented several noteworthy aspects that warrant attention.

At the interest dimension level, the machine-inferred interest scores show comparable Cronbach's αs with self-reported interest scores. In the training sample, the average Cronbach's αs across the eight dimensions for self-reported and machine-inferred scores are .86 and .87, respectively. In the test sample, the average Cronbach's αs for the two sets of scores are .84 and .87, respectively. The results showed that the machine-inferred scores have comparable Cronbach's αs as those of self-reported scores. Note that the average Cronbach's αs of machine-inferred scores for the training and the test sample are also comparable, indicating that the machine learning models I built have excellent cross-sample generalizability in terms of Cronbach's αs.

At the scale level, the self-reported interest scores demonstrated great internal consistency reliabilities and split-half reliabilities. Averaged Cronbach's αs across all self-reported scales for the training and test sample are .90 and .88, respectively. The averaged split-half reliabilities across all self-reported scales for the two samples are $\bar{r}$ = .87 and .86, respectively.

The split-half reliabilities for machine-inferred scores are somewhat lower than those of the self-reported scores in the training sample but are substantially lower than the self-reported scores in the testing sample. Averaged split-half reliabilities across all machine-inferred scales for the training and test sample are $\bar{r}$ = .60 and .50, respectively. Averaged split-half reliabilities for each dimension in the training and test sample are as follows: $\bar{r}$ =.67 and .57 for Health Science scales; $\bar{r}$ = .62 and .54 for Creative Expression scales; $\bar{r}$ = .69 and .55 for Technology scales; $\bar{r}$ = .62 and .49 for People scales; $\bar{r}$ = .56 and .51 for Organization scales; $\bar{r}$ = .53 and .42 for Influence scales; $\bar{r}$ = .63 and .44 for Nature scales; and $\bar{r}$ = .61 and .54 for Things scales.

The aforementioned results suggest encouraging evidence regarding the internal consistency of machine-inferred vocational interest dimension scores. However, the split-half reliabilities of machine-inferred interest scale scores were found to be lower when compared to self-reported scale scores. Additionally, although the split-half reliabilities of machine-inferred scale scores derived from the test sample are lower than those in the training sample, the Cronbach's αs of the machine-inferred scale scores in those two samples demonstrated reasonable generalizability across different samples.

Factorial Validity

The adequacy of the set-ESEM model fit was evaluated using various statistical measures, including the chi-square goodness-of-fit test, the Tucker–Lewis index (TLI) based on maximum likelihood estimation, the comparative fit index (CFI), the root-mean-squared error of approximation (RMSEA), and the standardized root-mean-squared residual (SRMR). For the training sample, the results were as follows: chi-square (χ^2) = 7962.46, degrees of freedom (*df*) = 2,616, $p < .01$, CFI = .82, TLI = .78, RMSEA = .09, SRMR = .04. For the test sample, the results were: χ^2 = 7511.01, *df* = 2,616, $p < .01$, CFI = .62, TLI = .52, RMSEA = .15, SRMR = .05. The model fit indices did not meet the widely used rule of thumb criteria proposed by Hu and Bentler (1999) for good model fit. However, the results obtained from the training sample could be considered acceptable given the complex structure of the model (i.e., SETPOINT model) I used here, and the model fit results reported in previous literature (e.g., Fan et al., 2023; Zhang et al., 2020). The results in the test sample are not acceptable.

Table 3 presents the factor loadings of the set-ESEM model for both the self-reported interest scores and the machine-inferred interest scores in the test sample. Table A3 in Appendix provided the the factor loadings for the training sample. The results indicate distinct patterns where the basic scales load onto their respective SETPOINT dimensions.

In the case of self-reported interest scores in the test sample, all basic scales demonstrated the highest loadings on their designated dimensions, with minimal cross-loadings on other dimensions. The sole exception was the Professional Advising scale, which was loaded the highest on People instead of Influence. Similarly, for the machine-inferred interest scores in the test sample, the overall patterns of the SETPOINT dimensions were clearly recovered. However, the Culinary Art scale loaded the highest on Technology rather than Creative Expression, the Humanities & Foreign Language scale had the highest loading on Nature instead of People, the Religious Activities had the highest loading on Influence instead of People, the Human Resources scale had the highest loading on Influence rather than Organization, and Finance had the highest loading on Influence rather than Organization. However, most of the loadings on the intended factors were still moderately high. Overall, the machine-inferred interest scores closely mirrored the patterns and structure observed in the self-reported interest scores.

Further results show that TCCs of the eight dimensions (i.e., Health Science, Creative Expression, Technology, People, Organization, Influence, Nature, Things) for the training sample were .93, .96, .92, .97, .91, .96, .88, .90, indicating that the factor structures of machine-inferred scores for Creative Expression, People, and Influence dimensions are virtually identical to those of corresponding self-reported scores, and those of machine-inferred scores for Health Science, Technology, Organization, Nature, and Things also have fair similar factor structures to the self-reported scores of their corresponding dimensions. The TCCs of the eight dimensions for the test sample are .87, .89, .53, .73, .82, .89, .74, .90, which are somewhat lower than the training sample. However, TCCs of half of the dimensions (i.e., Health Science, Creative

Expression, Influence, and Things) still indicate fair similarity of factor structures between self-report and machine-inferred interest scores.

In terms of magnitude similarity, RMSEs for the eight dimensions range from .07 to .13 in the training sample, which were small in general. In the test sample, RMSEs are .14 ~ .19 except for .26 for Technology dimension. In addition, the plots in Figure 1 visualize that differences of factor loadings are generally small between self-report and machine-inferred scores as most of the dots are close to the line Y = X.

In general, although the model fit indices are not ideal and there are several exceptions in factor loadings, the machine-inferred interest scores revealed factor structures that are similar to those under the self-report interest scores. Moreover, the TCCs and RMSEs indicate fair pattern similarity and magnitude similarity between interest scores based on the two measurement methods. However, when interpreting those results, note that the small sample size (n = 88) used in the test sample is probably not appropriate for set-ESEM analysis.

Convergent and Discriminant Validity

The latent and manifest vocational interest dimensional MTMM matrices were presented in Table 4 and Table 5. Table 6 presented the convergent and discrimination indices of latent level and manifest level in the training and test sample. At the latent variable level, C1s are .68 and .36 in training and test sample, respectively. C1s in the manifest level are .68 and .40, respectively. The C1s show that trait-specific variance and person main effects contribute large amount of observed variance, which indicate great convergent validity. However, C1s in the test sample at both latent and manifest level exhibit significantly lower values compared to those observed in the training sample.

D1s at the latent dimension level are .55 and .22 in the training and test sample, respectively. At the manifest level, they are .51 in the training sample and .24 in the test sample. As D1 indicates trait-specific variance, a contrast of C1 and D1 reveals that trait-specific variances contributed most of the convergence while person main effects contributed a small proportion of it. D2s at the latent dimension level are .49 and .20 in the training and test sample, respectively. D2s are somewhat lower at the manifest level (.36 and .06, respectively) in the training sample and test sample. However, all D2s are positive, indicating the variance accounted for by trait-specific variance is larger than the method-specific variance. D2as in the training and test sample are .50 and .21 at the latent dimension level and are .41 and .13 at the manifest level. As D2as were calculated using only machine-inferred monomethod correlations, higher D2a's compared to D2s indicate that machine learning approach yields lower levels of method variance. Results of D1, D2, and D2a indicate that the machine learning method has good discriminant validity. However, the convergence indices in the test sample, at both the latent and manifest levels, demonstrate notably lower values compared to those observed in the training sample.

In addition, the average absolute intercorrelations of self-reported interest dimension scores in the training and test sample are .21 and .17 at the latent dimension level and are .36 and .40 at the manifest level. However, the average absolute intercorrelations of machine-inferred interest scores in the training and test sample are somewhat lower than those of self-reported scores (latent level: .18 and .15; manifest level: .27 and .27). Such results further support the discriminant validity of machine learning method.

Moreover, as shown in Table 4, the latent dimension level discrimination indices are overall higher than those at the manifest level, indicating better discriminant validity of machine-inferred interest scores at the latent level than those at the manifest level. Table 4 also shows generally lower convergence and discrimination indices in the test sample compared to those in the training sample. Although this may indicate a lower level of cross-sample generalizability of convergent and discriminant validities, the much smaller sample size of the test sample (n = 88) may also account for lower convergence and discrimination indices in the test sample.

The findings above suggest that machine-inferred vocational interest latent dimension scores demonstrated excellent convergent and good discriminant validity. Machine-inferred vocational interest manifest dimension scores have comparable convergent validity but somewhat lower discriminant validity compared to those at the latent level, especially in the test sample. Finally, it is important to note that the lower levels of convergent and discriminant validity observed in the tests ample may be attributed to the small sample size.

Criterion-Related and Incremental Validity

Examining how interest congruence predicts external criteria is plausible as most participants in the test sample can be categorized into two majors: Marketing (n = 41) and Business (n = 15) after participants without GPA or high school GPA were removed. To obtain the congruence between student's interest profile and their major's interest profile, I calculated the average of all students' interest scores on each of the eight dimensions for each major and used this set of eight scores as the major's interest profile. Congruence was indexed by the correlation between the student's interest profile (i.e., interest scores on the eight dimensions) and the student's major's interest profile. Two sets of interest congruence indices were calculated such that each student has one interest congruence index that was calculated using the self-reported interest scores and another interest congruence index calculated using the machine-inferred interest scores.

The bivariate correlations between congruence indices calculated using self-reported interest dimension scores and the machine-inferred interest dimension scores and the six criterion variables are presented in Table 7. Partial correlations after controlling high school GPA were also calculated, although none of the correlations between self-reported or machine-inferred interest dimension congruence index and the criterion variables is significant (see Table 8). However, correlation coefficients between interest congruence based on machine-inferred dimension scores and the criterion variables (average of absolute correlations is .10) tend to be larger than those between interest congruence based on self-reported dimension scores and criterion variables (average of absolute correlations is .06).

Further hierarchical regression analysis examined the incremental validity of machine-inferred interest dimension scores (see Table 9). There are several noteworthy observations. First, high school GPA was a significant predictor of major satisfaction, withdrawal intention, and cumulative GPA. Second, after controlling for high school GPA, interest congruence based on self-reported dimension scores significantly predicts major satisfaction ($\beta = .44$, $p < .001$). Interestingly, after interest congruence based on machine-inferred dimension scores were entered into the model predicting absenteeism, the β of interest congruence based on self-reported dimension scores became more negative (β decreased from -.22 to -.32) and significant ($p < .05$). Third, interest congruence based on machine-inferred dimension scores did not explain any additional variance in predicting the six criteria. However, as I explained in previous sections, the small sample size ($n = 56$) may contribute to the overall insignificance of the results examined.

Overall, even though correlation and regression analysis did not provide significant support for the criterion validity and incremental validity of interest congruence based on machine-inferred dimension scores, the results are not discouraging considering the small sample sized used in these analyses.

Discussion

Major Findings and Implications

Recognizing the significance of vocational interests is crucial as it plays an important role in shaping career choices, academic engagement, and overall life satisfaction (Krumboltz et al., 1976; Stoll et al., 2017). Individuals with strong vocational interest experience a heightened sense of motivation and engagement, leading to increased focus and commitment towards their chosen field of study (Lent et al., 2002). However, the measurement of vocational interests relies heavily on self-report method, which is subject to factors such as social desirability bias and the lack of self-awareness of respondents, leading to potential inaccuracies in measurement (Gati et al., 1996; Hough et al., 2001). Completing self-report measures can also be arduous and time consuming, potentially resulting in respondents providing less thorough responses (Tracey, 2020). The current study aims to investigate the plausibility of using an AI-based chatbot to measure vocational interests. Specifically, I built machine learning models for each of the 41 basic interest scales in the SETPOINT model and examined the psychometric properties of the machine-inferred scores in terms of: (1) reliability, (2) factorial validity, (3) convergent and discriminant validity, and (4) criterion-related and incremental validity. I found overall positive evidence supporting the utility of ML approach in predicting vocational interest scores in the three of the validity (i.e., reliability, factorial validity, and convergent and discriminant validity) but unsatisfactory results regarding criterion-related and incremental validity.

Results showed that machine-inferred interests scores have comparable internal consistency in the dimension level compared to self-reported scores. However, the split-half reliabilities of machine-inferred basic interest scale scores did not reach the acceptable level in the test sample and are lower than those reported in other ML studies (e.g., Fan et al., 2023; Hoppe et al., 2018; Wang & Chen, 2020; Youyou et al., 2015), and the cross-sample generalizability of split-half reliability is not as good as reported in Fan et al. (2023). One possible reason for the overall modest split-half reliabilities for the machine-inferred scale scores is that the models predict the vocational interest scale scores better when more texts are included in the model. As we mentioned in the Analysis section, the split-half reliabilities were obtained by applying the model to half of each participant's chat text and calculating the correlation between the machine-inferred scores of the two halves for each scale. The small sample size of the test sample and that the training sample and that the test sample has distinct major distribution may also contribute to lower average split-half reliability than those reported in previous similar ML studies and lower cross-sample generalizability.

My results provide some new evidence supporting the discriminant validity of machine learning approach, considering that the machine learning approach has usually been found of having poorer discriminant validity than self-reported scores (Fan et al., 2023; Harrison et al., 2019; Hickman et al., 2022; Marinucci et al., 2018; Park et al., 2015). Comparing the latent level discrimination indices in the training sample of this study with those reported in similar personality ML studies (see Table 6), several observations are noteworthy. First, D1 and D2 are overall higher in this study, indicating (1) the proportion of trait-specific variance is larger than the person-specific variance, and this tendency is higher in this study compared to other listed study, and (2) the proportion of trait variance is higher than the method variance, and the tendency is also higher in this study. Second, a slightly higher D2a than D2 in our study indicates that the machine method yielded less proportion of method variance compared to the self-report method, a sign of better discriminant validity (machine method vs. self-report method) that is highlighted in this study. In addition, machine-inferred dimension scores have overall lower levels of average absolute intercorrelations compared to those of self-reported scores, which also supports a better discriminant validity of machine-inferred scores compared to self-reported scores. Finally, the method variance of machine-inferred scores is overall lower than those of self-reported scores, which provided unique evidence for better discriminant validity of machine method than self-report method, contrasting to the opposite patterns in Fan et al. (2023) and Park et al. (2015). Considering the difference of study objects (vocational interest vs. personality) in this study and other mentioned ML studies, I offer one potential explanation for such findings in this study. It is possible that vocational interests are more concrete conceptions compared to personality as perceived by respondents, which made it easier for them to clarify and differentiate their inner reflections on various vocational interests compared to personality traits when completing the questionnaire. When chatting with the chatbot in this study, respondents were asked a lot of questions about their preferred work activities. To answer such questions, respondents only need to focus on the fact (i.e., activities) and describe it. However, in personality measures, the items/questions asked are more abstract and thus require more thinking and are less easy for respondents to differentiate between traits.

The results on criterion-related and incremental validity are not as encouraging, however. None of the partial correlations between interest congruence indices and the criterion variables nor the hierarchical regression results are significant. However, interest congruence derived from machine-inferred scores have overall larger absolute partial correlations with criterion variables ($\overline{r}$ = .10) compared to those based on interest congruence derived from self-reported scores ($\overline{r}$ = .06). Despite the small sample size (*n* = 56) used for the analysis of partial correlation and hierarchical regression, the distinct major composition of training sample and test sample may also contribute to insignificant results. Participants with different majors may have distinct vocational interests. The majors in the training sample may not be comprehensive enough so models built based on the training sample may not be able to predict the vocational interests scores of participants in the test sample very well.

Contributions and Strengths

This study is the first systematic examination, to my best knowledge, of the plausibility of using AI chatbot in predicting vocational interests. Moreover, as there is significant lack of attention on individual differences except for personality in the ML literature, the current study represents a further step in discovering the potential of AI chatbot and ML approach in predicting various individual differences.

Also, the study found good reliability, factorial validity, and convergent and discriminant validity of ML approach in predicting vocational interests, which opened a new window for another effective way of measuring vocational interest in addition to the traditional self-report measures. As we mentioned, self-report vocational interest questionnaires can be long and arduous to complete (Tracey, 2020), and are subject to social desirability bias and lack of self-awareness of respondents (Gati et al., 1996; Hough et al., 2001). The comprehensive evidence of the psychometric properties established the foundation for future research and application of ML approach in measuring vocational interests.

The favorable findings regarding factorial validity observed in the present study make a valuable contribution to the existing vocational interest literature as it examined the most recent and comprehensive vocational interest model, the SETPOINT model. The findings demonstrate that not only self-reported basic interest scale scores align well with the SETPOINT model, the machine-inferred basic interest scale scores also recover this structure well. The results in the training sample of the current study provided supportive evidence for the use of SETPOINT model in research.

Several methodological strengths are also noteworthy in the present study. First, following Fan et al. (2023), I built ML models at the level of basic interest scale, which enables the examination of internal consistency at the dimension level, factorial validity, and convergent and discriminant validity at the latent level. Such detailed investigation cannot be achieved if models were built at the dimension level. In addition, the present study design includes a training sample and an independent test sample, which facilitates the investigation of the extent to which the psychometric properties of the measures can be generalized across different samples.

Limitations

One of the most significant limitations of this study is regard to the samples used for analysis. First, the sample sizes for both the training sample and test sample are small, especially for the test sample. This issue is more salient considering that the number of predictors in the machine learning models are huge, which may lead to the building of unstable models. However, the psychometric properties of the interest scores obtained in the training sample in the current study are overall very promising, which encourages future research to investigate those properties further using larger sample size. Second, the major composition of participants in the training and test sample are distinct. The participants in the training sample recruited from SONA system have various majors such as Psychology, Kinesiology, Computer Science, and Health Science, etc., while participants in the test sample are business school students whose majors are mostly Business and Marketing. Unlike personality that is more abstract and general, vocational interests are more concrete and sensitive to the majors of participants chose, which might have led to the difference of vocational interests profile in the training and the test sample. The different major compositions in the training sample and test sample may contribute to the relatively modest split-half reliabilities, discriminant validity, factorial validity, and the insignificant criterion-related and incremental validity in the test sample, especially when the sample size of test sample is small. Future studies should incorporate larger sample sizes in both the training and test sample, and recruit participants from various majors to increase the representativeness of the ML models built based on the training sample.

Moreover, the results did not support the criterion-related and incremental validity of interest congruence based on machine-inferred vocational scores. The limitation of the samples mentioned above could be an important contributing factor to such results. Another possible reason is that the way I calculated the major interest profiles did not fully reflect the interest profile of the major. Using the average of students' interest profiles as the major's interest profile is the "incumbent approach" from Holland (1997). However, individual profiles within the major may not agree strongly if many students chose the major for reasons other than interest in the major. Another way to determine the major interest profile is referring to interests information of occupations listed on O*NET (e.g., Nye et al., 2017). However, this is hard to achieve in the current study which adopted the most recent SETPOINT model while interests information listed on O*NET is based on the RIASEC model. Future research may explore other ways of determining the major interest profile when SETPOINT model is being used.

Last, although I speculated that interaction with AI-based chatbot brings better experience for respondents, the current study did not examine participants' perceptions of the AI-based chatbot and the ML approach. Although interacting with a human-like chatbot can be fun and less tedious for respondents, realizing that their responses will be analyzed by ML models rather than human agents may evoke negative reactions (Gonzalez et al., 2019; Langer et al., 2019). Gonzalez et al. (2022) suggests that applicants may perceive less procedural justice in ML-based assessments because they may have less opportunities to voice their opinions than when they are evaluated by human judgment. Further, given that vocational interests are measured in both counseling context and in high-stake selection context, respondents' reactions may also be different in the two contexts. Future research should explicitly examine respondents' reactions toward AI-based chatbot, the ML approach, and the combination of these two in measuring vocational interests.

Future Research Directions

The results of the study encourage several other lines of future research. First, future research could consider conducting longitudinal studies which allow for the examination of change and stability of vocational interest scores over time. One of the advantages of the ML approach is its ability to handle data from a sizable sample and regularly incorporate new data to update the model parameters. (Gonzalez et al., 2022), and it is also possible to periodically receive new input from respondents. Such features would enable researchers to easily obtain data of respondents' vocational interests across a period of time and examine the pattern of changing of vocational interests over time.

Second, the psychometric properties of machine-inferred vocational interests scores derived from AI-based chatbot method should be examined in selection contexts. The promising results found in the current study were based on undergraduate sample for research purposes. It is not clear whether encouraging results will be revealed in applicant sample under selection contexts. Future study should examine if the findings are still valid and if machine-inferred vocational interests scores could predict work-related outcomes such as performance, turnover, job satisfaction, organizational citizenship behavior, and counterproductive work behavior, etc.

Moreover, future research should examine if machine-inferred vocational interests scores are resistant to faking. While one view argues that machine-inferred personality scores cannot be faked due to the use of numerous meaningless features, others content that individuals may still engage in impression management that could captured by the algorithm, which could influence the scores (Fan et al., 2023). Vocational interest, being a more concrete concept than personality, might be more prone to faking in self-reported measures. Therefore, it is worthwhile to empirically investigate whether machine-generated vocational interest scores are more resistant to faking compared to self-report methods.

References

Assouline, M., & Meir, E. I. (1987). Meta-analysis of the relationship between congruence and well-being. measures. *Journal of Vocational Behavior*, *31*(3), 319-332. https://doi.org/10.1016/0001-8791(87)90046-7

Azucar, D., Marengo, D., & Settanni, M. (2018). Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis. *Personality and individual differences, 124*, 150-159. https://doi.org/10.1016/j.paid.2017.12.018

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. *Personnel psychology*, *44*(1), 1-26. https://doi.org/10.1111/j.1744-6570.1991.tb00688.x

Bleidorn, W., & Hopwood, C. J. (2019). Using machine learning to advance personality assessment and theory. *Personality and Social Psychology Review*, *23*, 190-203. https://doi.org/10.1177/1088868318772990

Cer, D., Yang, Y., Kong, S.-Y., Hua, N., Limtiaco, N., John, R., Constant, N., Guajardo Cespedes, M., Yuan, S., Tar, C., Sung, Y.-H., Strope, B., & Kurzweil, R. (2018). Universal Sentence Encoder. https://doi.org/10.48550/arXiv.1803.11175

Chernyshenko, O. S., Stark, S., & Drasgow, F. (2011). Individual differences: Their measurement and validity. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology, Vol. 2. Selecting and developing members for the organization* (pp. 117–151). American Psychological Association. https://doi.org/10.1037/12170-004

Chernyshenko, O. S., Stark, S., & Nye., C. D. (2019). Interest measurement. In C. D. Nye & J. Rounds (Eds.), *Vocational interests: Rethinking their role in understanding workplace behavior and practice, SIOP organizational frontiers series* (pp. 80–96). Routledge. https://doi.org/10.4324/9781315678924-5

Deci, E., & Ryan, R. (2012). Self-determination theory. In P. A. Van Lange A. W. Kruglanski, & E. T. Higgins (Ed.), *Handbook of theories of social psychology* (Vol. 1, pp. 416-437). SAGE Publications, Ltd. https://dx.doi.org/10.4135/9781446249215.n21

Donnay, D. A. C., Morris, M. L., Schaubhut, N. A., & Thompson R. C. (2005). *StrongInterest Inventory® manual*. Sunnyvale, CA: The Myers-Briggs Company.

Fan, J., Sun, T., Liu, J., Zhao, T., Glorioso, M., Chen, Z., Zhang, B., Hack., E. (2023). How well can an AI chatbot infer personality? Examining psychometric properties of machine-inferred personality scores. *Journal of Applied Psychology.* Advance online publication. https://dx.doi.org/10.1037/apl0001082

Gati, I., Krausz, M., & Osipow, S. H. (1996). A taxonomy of difficulties in career decision making. *Journal of Counseling Psychology, 43*(4), 510–526. https://doi.org/10.1037/0022-0167.43.4.510

Gati, I., & Kulcsár, V. (2021). Making better career decisions: From challenges to opportunities. *Journal of Vocational Behavior*, *126*, 103545. https://doi.org/10.1016/j.jvb.2021.103545

Gonzalez, M. F., Capman, J. F., Oswald, F. L., Theys, E. R., & Tomczak, D. L. (2019). Where's the I-O"? Artificial intelligence and machine learning in talent management systems. *Personnel Assessment and Decisions, 5*(3), 33–44. https://doi.org/10.25035/pad.2019.03.005

Gonzalez, M. F., Liu, W., Shirase, L., Tomczak, D. L., Lobbe, C. E., Justenhoven, R., & Martin, N. R. (2022). Allying with AI? Reactions toward human-based, AI/ML-based, and augmented hiring processes. Computers in Human Behavior, 130, 107179. https://doi.org/10.1016/j.chb.2022.107179

Gottfredson, L. S. (2005). Applying Gottfredson's Theory of Circumscription and Compromise in Career Guidance and Counseling. In S. D. Brown & R. W. Lent (Eds.), *Career development and counseling:Putting theory and research towork* (pp. 71–100). John Wiley & Sons, Inc..

Gupta, S., Jagannath, K., Aggarwal, N., Sridar, R., Wilde, S., & Chen, Y. (2019). Artificially Intelligent (AI) Tutors in the Classroom: A Need Assessment Study of Designing Chatbots to Support Student Learning. *PACIS 2019 Proceedings*, 213. https://aisel.aisnet.org/pacis2019/213

Hanna, A., & Rounds, J. (2020). How accurate are interest inventories? A quantitative review of career choice hit rates. *Psychological Bulletin*, *146*(9), 765–796. https://doi.org/10.1037/bul0000269

Hansen, J. I. C. (2019). Interest inventories. In G. Goldstein, D. N. Allen, & J. Deluca (Eds.), *Handbook of psychological assessment* (pp. 169-190). Cambridge, MA: Academic Press. http://dx.doi.org/10.1016/B9780-12-802203-0.00006-7

Harrison, J. S., Thurgood, G. R., Boivie, S., & Pfarrer, M. D. (2019). Measuring CEO personality: Developing, validating, and testing a linguistic tool. Strategic Management Journal, 40, 1316–1330. https:// doi.org/10.1002/smj.3023

Hickman, L., Bosch, N., Ng, V., Saef, R., Tay, L., & Woo, S. E. (2022). Automated video interview personality assessments: Reliability, validity, and generalizability investigations. *Journalof Applied Psychology*, *107*(8), 1323–1351. https://doi.org/10.1037/apl0000695

Holland, J. L. (1966). *The Psychology of Vocational Choice: A Theory of Personality Types and Model Environments*. Blaisdell.

Holland, J. L. (1985). *Vocational Preference Inventory*. Lutz, FL: Psychological Assessment Resources.

Holland, J. L. (1997). *Making vocational choices* (3rd ed.). Odessa, FL: Psychological Assessment Resources.

Holland, J. L., Fritzsche, B., & Powell, A. (1994). *The self-directed search: Technical manual*. Odessa, FL: Psychological Assessment Resources.

Hoppe, S., Loetscher, T., Morey, S. A., & Bulling, A. (2018). Eye movements during everyday behavior predict personality traits. *Frontiers in Human Neuroscience, 12*, Article105. https://doi.org/10.3389/fnhum.2018.00105

Hough, L. M., Barge, B., & Kamp, J. (2001). Assessment of personality, temperament, vocational interests, and work outcome preferences. In J. P. Campbell & D. J. Knapp (Eds.), *Exploring the limits of personnel selection and classification* (pp. 111–154). Mahwah, NJ: Erlbaum.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling, 6(1), 1–55. https://doi.org/10.1080/10705519909540118

IBM Cloud Education. (2020). *Machine Learning*. https://www.ibm.com/cloud/learn/machine-learning.

Ingerick, M., & Rumsey, M. G. (2014). Taking the measure of work interests: Past, present, and future. *Military Psychology, 26*, 165–181. http://dx.doi.org/10.1037/mil0000045

Kern, M. L., McCarthy, P. X., Chakrabarty, D., & Rizoiu, M. A. (2019). Social media-predicted personality traits and values can help match people to their ideal jobs. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(52), 26459–26464. Advance online publication. https://doi.org/10.1073/pnas.1917942116

Kim, S., Lee, J., & Gweon, G. (2019, May). Comparing data from chatbot and web surveys: Effects of platform and conversational style on survey response quality. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1–12). https://doi.org/10.1145/3290605.3300316

Kirkendall, C. D., Nye, C. D., Rounds, J., Drasgow, F., Chernyshenko, O. S., & Stark, S. (2020). Adaptive vocational interest diagnostic: Informing and improving the job assignment process. Military Psychology, 32(1), 91–100. https://doi.org/10.1080/08995605.2019.1652480

Kiselev, P., Kiselev, B., Matsuta, V., Feshchenko, A., Bogdanovskaya, I., & Kosheleva, A. (2020). Career guidancebased on machine learning: social networks in professional identity construction. *Procedia Computer Science*, *169*, 158-163. https://doi.org/10.1016/j.procs.2020.02.128

Kosinski, M., Stillwell, D., & Graepel, T. (2013).Private traits and attributes are predictable from digital records of human behavior. *Proceedingsof the National Academy of Sciences, 110*, 5802 - 5805. https://doi.org/10.1073/pnas.1218772110

Krumboltz, J. D., Mitchell, A. M., & Jones, G. B. (1976). A social learning theory of career selection. *The Counseling Psychologist*, *6*(1), 71-81. https://doi.org/10.1177/001100007600600117

Kuder, F. (1983). *Manual, Kuder preference record: Form C*. Adel, IA: National Career Assessment Services.

Langer, M., König, C. J., & Papathanasiou, M. (2019). Highly automated job interviews: Acceptance under the influence of stakes. *International Journal of Selection and Assessment*, *27*(3), 217–234. https://doi.org/10.1111/ijsa.12246

Lee, T., Jagannath, K., Aggarwal, N., Sridar, R., Wilde, S., Hill, T., & Chen, Y. (2019). Intelligent career advisers in your pocket? A need assessment study of chatbots for student career advising. *AMCIS 2019 Proceedings*, 14. https://aisel.aisnet.org/amcis2019/human_computer_interact/human_computer_interact/14

Lent, R. W., Brown, S. D., & Hackett, G. (2002). Social cognitive career theory. In D. Brown (Ed.), *Career choice and development* (4th ed., pp. 255-311). Jossey-Bass.

Leutner, K., Liff, J., Zuloaga, L., & Mondragon, N. (2021). Hirevue's assessment science [White paper]. https://Hirevue.com.https://webapi. hirevue.com/wp-content/uploads/2021/03/HireVue-Assessment-Sciencewhitepaper-2021.pdf

Lorenzo-Seva, U., & ten Berge, J. M. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 2(2), 57–64. https://doi.org/10.1027/1614-2241.2.2.57

Marinucci, A., Kraska, J., & Costello, S. (2018). Recreating the relationship between subjective. wellbeing and personality using machine learning: An investigation into Facebook online behaviours. *Big Data and Cognitive Computing*, *2*(3), Article 29. https://doi.org/10.3390/bdcc2030029

Marsh, H. W., Guo, J., Dicke, T., Parker, P. D., & Craven, R. G. (2020). Confirmatory factor analysis (CFA), exploratory structural equation modeling (ESEM), and Set-ESEM: Optimal balance between goodness of git and parsimony. *Multivariate Behavioral Research*, *55*, 102-111. https://doi.org/10.1080/00273171.2019.1602503

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437–455. https://doi.org/10.1037/a0028085

Nauta, M. M. (2007). Assessing college students' satisfaction with their academic majors. Journal of Career Assessment, 15, 446–462. https://doi.org/10.1177/1069072707305762

Nye, C., & Rounds, J. (Eds.). (2019). Vocational Interests in the Workplace: Rethinking Behavior at Work (1st ed.). Routledge. https://doi.org/10.4324/9781315678924

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*, 384–403. https://doi.org/10.1177/1745691612449021

Nye, C. D., Su, R., Rounds, J., & Drasgow, F. (2017). Interest congruence and performance: Revisiting recent meta-analytic findings. *Journal of Vocational Behavior*, *98*, 138-151. https://doi.org/10.1016/j.jvb.2016.11.002

Oswald, F. L., Behrend, T. S., Putka, D. J., & Sinar, E. (2020). Big data in industrial-organizational psychology and human resource management: forward progress for organizational research and practice. *Annual Review of Organizational Psychology and Organizational Behavior*, *7*, 505-533. https://doi.org/10.1146/annurev-orgpsych-032117-104553

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., Ungar, L. H., & Seligman, M. E. (2015). Automatic personality assessment through social media language. *Journal of Personality and Social Psychology, 108*, 934–952. https://doi.org/10.1037/pspp0000020

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofter, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D. Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

Rounds, J., Hoff, K., Lewis, P. (2021). *O*NET® Interest Profiler Manual.* National Center for O*NET Resource. Center. Retrieved from: https://www.onetcenter.org/dl_files/IP_Manual.pdf

Rounds, J., & Su, R. (2014). The nature and power of interests. *Current Directions in Psychological Science, 23*, 98 –103. http://dx.doi.org/10 .1177/0963721414522812

Schmitt, N., Oswald, F. L., Friede, A., Imus, A., & Merritt, S. (2008). Perceived fit with an academic environment: Attitudinal and behavioral outcomes. J*ournal of Vocational Behavior, 72*(3), 317-335. https://doi.org/10.1016/j.jvb.2007.10.007

Schwartz, S. H. (1992). Universals in the content and structure of values: Theory and empirical tests in 20 cultural groups. In M. Zanna (Ed.), *Advances in experimental social psychology* (*Vol. 25*, pp. 1-65). Academic Press. https://doi.org/10.1016/S00652601(08)60281-6

Schwartz, S. H. (2017). The refined theory of basic values. In S. Roccas & L. Sagiv (Eds.), *Values and behavior: Taking a cross cultural perspective* (pp. 51–72). Springer International Publishing. https://doi.org/10.1007/978-3-319-56352-7_3

Settanni, M., Azucar, D., & Marengo, D. (2018). Predicting individual characteristics from digital traces on social media: A meta-analysis. *Cyber Psychology, Behavior & Social Networking*, *21*, 217–228. https://doi.org/10.1089/cyber.2017.0384

Song, Q. C., Shin, H. J., Tang, C., Hanna, A., & Behrend, T. (2022). Investigating machine learning's capacity to enhance the prediction of career choices. *Personnel Psychology*. https://doi.org/10.1111/peps.12529

Spector, P. E., & Jex, S. M. (1998). Development of Four Self-Report Measures of Job Stressors and Strain: Interpersonal Conflict at Work Scale, Organizational Constraints Scale, Quantitative Workload Inventory, and Physical Symptoms Inventory. *Journal of Occupational Health Psychology, 3*, 356-367. https://doi.org/10.1037/1076-8998.3.4.356

Staudinger, U. M., Fleeson, W., & Baltes, P. B. (1999). Predictors of subjective physical health and global well-being: Similarities and differences between the United States and Germany. *Journal of Personality and Social Psychology*, *76*(2), 305. https://doi.org/10.1037/0022-3514.76.2.305

Stoll, G., Rieger, S., Lüdtke, O., Nagengast, B., Trautwein, U., & Roberts, B. W. (2017). Vocational interests assessed at the end of high school predict life outcomes assessed 10 years later over and above IQ and Big Five personality traits. *Journal of personality and social psychology*, *113*(1), 167. http://dx.doi.org/10.1037/pspp0000117

Stoughton, J. W., Thompson, L. F., & Meade, A. W. (2015). Examining applicant reactions to the use of social networking websites in pre-employment screening. *Journal of Business and Psychology, 30*, 73-88. https://doi.org/10.1007/s10869-013-9333-6

Strong Jr., E. K. (1943). *Vocational interests of men and women*. Palo Alto, CA: Stanford University Press.

Su, R., Tay, L., Liao, H.-Y., Zhang, Q., & Rounds, J. (2019). Toward a dimensional model of vocational. Interests. *Journal of Applied Psychology, 104*(5), 690–714. http://dx.doi.org/10.1037/apl0000373

w

Super, D. E. (1953). A theory of vocational development. *American Psychologist*, *8*, 185–190. https://doi.org/10.1037/h0056046

Tay, L., Woo, S. E., Hickman, L., & Saef, R. M. (2020). Psychometric and Validity Issues in Machine Learning Approaches to Personality Assessment: A Focus on Social Media Text Mining. *European Journal of Personality, 34*, 826–844. https://doi.org/10.1002/per.2290

Tracey, T. J. G. (2002). Personal Globe Inventory: Measurement of the spherical model of interests and competence beliefs. *Journalof Vocational Behavior, 60*(1), 113–172. https://doi.org/10.1006/jvbe.2001.1817

Tracey, T. J. G. (2020). We can do that? Technological advances in interest assessment. *Journalof Career Assessment, 28*(1), 3–13. https://doi.org/10.1177/1069072719879910

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

Tsabari, O., Tziner, A., & Meir, E. I. (2005). Updated meta-analysis on the relationship between congruence and satisfaction. *Journal of Career Assessment, 13*, 216–232. http://dx.doi.org/10.1177/1069072704273165

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. *Journalof Applied Psychology, 96*(6), 1167–1194. https://doi.org/10.1037/a0024343

Wang, S., & Chen, X. (2020). Recognizing CEO personality and its impact on business performance: Mining linguistic cues from social media. Information & Management, 57(5), Article 103173. https://doi.org/10 .1016/j.im.2019.103173

Woehr, D. J., Putka, D. J., & Bowler, M. C. (2012). An examination of g-theory methods for modeling multitrait-multimethod data: Clarifying links to construct validity and confirmatory factor analysis. *Organizational Research Methods*, *15*(1), 134–161. https://doi.org/10.1177/1094428111408616

Xiao, Z., Zhou, M. X., Liao, Q. V., Mark, G., Chi, C., Chen, W., & Yang, H. (2020). Tell me about yourself: Using an AI-powered chatbot to conduct conversational surveys with open-ended questions. *ACM Transactions on Computer-Human Interaction, 27*(3), Article 15, 1-37. https://doi.org/10.1145/3381804

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences, 112*, 1036–1040. https://doi.org/10.1073/pnas.1418680112

Zickar, M. J., & Min, H. (2019). A history of vocational interest measurement. In C. D. Nye & J. Rounds (Eds.), *Vocational interests: Rethinking their role in understanding workplace behavior and practice* (pp. 59 79). New York, NY: Routledge. http://dx.doi.org/10.4324/9781 315678924-4

Zhang, B., Luo, J., Chen, Y., Roberts, B., & Drasgow, F. (2020). The road less traveled: A cross-cultural study of the negative wording factor in multidimensional scales. PsyArXiv. https://doi.org/10.31234/osf.io/2psyq

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B, Statistical Methodology, 67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

Tables and Figure

Table 1
Reliabilities of Self-Reported and Machine-Inferred Interest Scale Scores

	Training sample			Test Sample		
	SR scores (n = 236)		MA scores (n = 229)	SR scores (n = 98)		MA scores (n = 88)
Interest scales	Coefficient α	Split-half reliability	Split-half reliability	Coefficient α	Split-half reliability	Split-half reliability
Health Science (average)	.93	.90	.67	.89	.89	.57
Life Science	.91	.88	.65	.85	.86	.55
Medical Science	.94	.92	.69	.91	.90	.56
Health Care Service	.92	.91	.68	.89	.89	.60
Creative Expression (average)	.90	.86	.62	.88	.83	.54
Media	.87	.85	.52	.86	.84	.56
Applied Arts & Design	.84	.76	.66	.83	.75	.49
Music	.92	.86	.68	.88	.77	.53
Visual Arts	.93	.90	.59	.94	.90	.40
Performing Arts	.92	.88	.62	.87	.85	.65
Creative Writing	.94	.92	.61	.92	.89	.57
Culinary Art	.90	.88	.65	.85	.85	.58
Technology (average)	.91	.89	.69	.87	.82	.55
Engineering	.89	.84	.74	.85	.81	.60
Physical Science	.87	.86	.64	.83	.81	.44
Information Technology	.96	.94	.73	.93	.93	.55
Mathematics/Statistics	.93	.92	.65	.88	.73	.62
People (average)	.87	.82	.62	.86	.81	.49
Social Science	.87	.80	.60	.83	.76	.45
Humanities & Foreign Language	.88	.84	.59	.88	.83	.43
Teaching/Education	.87	.84	.59	.88	.75	.40
Social Service	.83	.79	.66	.79	.81	.67
Religious Activities	.90	.84	.66	.92	.88	.50
Organization (average)	.89	.86	.56	.85	.88	.51
Human Resources	.90	.84	.51	.86	.85	.42
Personal Service	.84	.76	.64	.85	.85	.63
Accounting	.93	.91	.49	.82	.93	.44
Office Work	.87	.88	.61	.91	.91	.57
Finance	.91	.89	.55	.83	.89	.48
Influence (average)	.90	.88	.53	.90	.89	.42
Management/Administration	.84	.86	.53	.90	.87	.37
Business Initiatives	.91	.88	.50	.87	.87	.49
Marketing/Advertising	.89	.84	.40	.90	.89	.34
Professional Advising	.89	.86	.54	.88	.80	.38

Public Speaking	.90	.89	.54	.92	.94	.43
Sales	.92	.89	.43	.90	.93	.42
Politics	.92	.91	.62	.89	.90	.43
Law	.93	.91	.63	.93	.94	.51
Nature (average)	.87	.84	.63	.87	.87	.44
Agriculture	.86	.88	.63	.87	.86	.39
Outdoors	.87	.83	.59	.87	.86	.44
Animal Service	.88	.83	.66	.88	.89	.48
Things (average)	.89	.88	.61	.91	.90	.54
Mechanics/Electronics	.91	.91	.71	.93	.95	.58
Transportation/Machine Operation	.86	.87	.64	.90	.90	.54
Construction/Woodwork	.91	.90	.68	.93	.89	.61
Physical/Manual Labor	.89	.88	.71	.89	.92	.56
Athletics	.93	.90	.52	.92	.92	.44
Protective Service	.86	.81	.43	.88	.85	.53

Note. SR scores = Self-reported interest scores; MA scores = Machine-inferred interest scores. "average" means dimension reliabilities were calculated by averaging the reliabilities of all the scales under the dimension. The sample size for self-reported scores and machine-inferred scores are different because some participants only have one of the two sets of scores.

Table 2

Reliabilities (Cronbach's α*) of Self-Reported and Machine-Inferred Interest Dimension Scores*

	Training sample (n = 229)		Test Sample (n = 88)	
Interest dimensions	SR scores	MA scores	SR scores	MA scores
Health Science	0.90	0.94	0.85	0.94
Creative Expression	0.89	0.91	0.88	0.92
Technology	0.85	0.92	0.77	0.91
People	0.80	0.80	0.78	0.80
Organization	0.88	0.79	0.87	0.80
Influence	0.91	0.92	0.90	0.91
Nature	0.81	0.85	0.83	0.86
Things	0.81	0.81	0.82	0.86

Note. SR scores = Self-reported interest scores; MA scores = Machine-inferred interest scores. Cronbach's αs were calculated using basic scale scores as item score

Table 3

Rotated Factor Loading Matrices Using Set-Exploratory Structural Equation Model Derived from the Test Sample

	SR scores								MA scores							
Basic interest scales	S	E	TE	P	O	I	N	TH	S	E	TE	P	O	I	N	TH
Life Science (S1)	**.68**	-.03	.27	-.04	-.05	.03	.18	.10	**.85**	.02	.35	-.04	-.05	-.02	.13	-.05
Medical Science (S2)	**.79**	-.11	.24	-.01	.04	.02	.01	.00	**.87**	-.10	.25	.01	-.09	-.03	.09	-.10
Health Care Service (S3)	**.80**	.01	-.15	.33	.10	-.05	-.07	.01	**.92**	-.22	-.12	.15	.01	-.01	-.01	-.06
Media (E1)	-.04	**.84**	.14	.12	-.18	.24	-.07	-.03	-.16	**.93**	-.08	.07	-.20	.26	-.08	.05
Applied Arits & Design (E2)	-.16	**.65**	-.16	.00	.27	-.09	.38	.04	-.03	**.92**	.11	-.02	.11	-.08	.15	.02
Music (E3)	.04	**.78**	.00	.13	-.06	-.05	-.03	.15	-.01	**.83**	.35	.20	-.13	.04	.00	.02
Visual Arts (E4)	-.05	**.64**	-.24	.08	.09	-.22	.31	-.02	.09	**.85**	-.10	-.17	-.01	-.14	.20	-.20
Performing Arts (E5)	-.10	**.76**	.12	.16	-.11	.14	-.22	.05	-.13	**.77**	-.17	.21	-.19	.14	.02	-.02
Creative Writing (E6)	.07	**.94**	.12	-.09	-.11	.06	-.29	-.03	-.07	**.84**	-.15	.12	-.18	.22	.08	-.06
Culinary Art (E7)	.08	**.20**	-.19	-.19	.24	.05	.31	.04	-.01	**.23**	.48	.38	.47	-.18	.30	-.37
Engineering (TE1)	.27	-.04	**.09**	-.10	-.03	.30	.26	.41	.05	-.04	**.61**	.04	.17	.13	.07	.44
Physical Science (TE2)	.54	.11	**.38**	-.26	.00	.06	.29	-.04	.51	.20	**.55**	-.08	-.07	.05	.24	.20
Information Technology (TE3)	.12	.11	**.35**	-.12	.30	-.04	-.06	.42	-.02	-.02	**.75**	-.02	.14	.01	-.12	.27
Mathematics/Statistics (TE4)	.15	-.07	**.41**	.15	.38	-.25	-.28	.28	.20	.00	**.74**	-.07	.09	-.04	-.22	.28
Social Science (P1)	.29	.43	.27	**.18**	.15	-.13	.21	-.08	.14	.33	.07	**.58**	-.05	.25	.25	-.16
Humanities & Foreign Language (P2)	.03	.30	.31	**.19**	-.11	.08	.23	.18	.01	.32	.13	**.46**	-.14	.19	.56	.04
Teaching/Education (P3)	.14	.17	-.12	**.44**	.04	.29	-.06	.15	.00	.02	-.03	**.90**	.19	-.05	-.12	.09
Social Service (P4)	.28	.18	-.30	**.44**	.27	-.01	.11	-.05	.20	.10	-.14	**.77**	.28	-.01	-.10	-.19
Religious Activities (P5)	.07	.20	-.01	**.41**	.14	.28	-.09	.06	.20	-.30	-.41	**.00**	-.01	.48	.17	-.17
Human Resources (O1)	-.02	-.02	.14	.02	**.78**	.34	.10	-.10	-.09	-.06	-.11	.19	**.59**	.49	.01	.06
Personal Service (O2)	.00	-.04	-.20	.48	**.50**	.18	-.10	.02	-.02	.12	-.30	.58	**.53**	.16	-.08	.01
Accounting (O3)	.11	-.05	.23	.00	**.84**	-.09	-.10	.00	-.04	-.40	.25	-.08	**.65**	.29	.05	-.01
Office Work (O4)	-.09	.06	.08	.08	**.65**	.14	.13	.09	-.24	.02	.20	.12	**.63**	.14	-.05	.18
Finance (O5)	.05	-.09	.06	-.29	**.51**	.42	-.12	.08	-.03	-.20	.32	-.21	**.58**	.43	.04	.06
Management/Administration (I1)	.01	-.14	.14	.16	.12	**.79**	.13	-.04	-.05	.00	.14	.04	.27	**.68**	-.30	.21

Business Initiatives (I2)	.18	-.09	-.09	-.22	.25	**.65**	-.09	.10	-.10	.12	.01	-.27	.56	**.53**	-.07	.17
Marketing/Advertising (I3)	-.19	.26	-.33	-.29	.13	**.66**	-.18	.01	-.03	.44	.01	-.29	.49	**.56**	-.11	.01
Professional Advising (I4)	.02	-.05	-.11	.33	.29	**.57**	-.13	.11	.09	.02	-.14	.49	.31	**.52**	-.12	.09
Public Speaking (I5)	-.05	.11	.12	.26	-.10	**.89**	.05	-.11	.05	.27	.20	.29	-.04	**.78**	-.13	-.10
Sales (I6)	.05	.08	-.23	-.32	.20	**.76**	-.21	.04	-.12	.15	.03	-.23	.59	**.54**	-.09	.06
Politics (I7)	-.06	.03	.31	.23	.02	**.66**	.29	-.05	-.06	-.40	-.06	.02	-.23	**.83**	.38	-.09
Law (I8)	.13	.05	.08	.17	.12	**.41**	.18	.01	.00	.04	-.06	.13	-.31	**.73**	.04	-.14
Agriculture (N1)	.12	-.04	.17	.01	.01	.05	**.50**	.48	.17	.08	.04	.02	.03	.02	**.71**	.39
Outdoors (N2)	.14	.08	.00	.01	-.07	.02	**.61**	.33	.06	.34	-.10	.17	.08	-.18	**.61**	.43
Animal Service (N3)	.05	.25	-.20	.12	.03	-.05	**.49**	.09	.46	.30	-.35	.26	.06	-.16	**.36**	.11
Mechanics/Electronics (TH1)	.06	-.01	.23	-.07	-.08	.13	.13	**.73**	.03	-.14	.32	-.07	-.05	.03	.22	**.73**
Transportation/Machine Operation (TH2)	-.04	.02	.03	.11	.04	-.08	.07	**.92**	-.09	-.19	.17	-.03	.01	-.10	.27	**.78**
Construction/Woodwork (TH3)	-.12	.13	.04	-.08	.01	-.06	.27	**.64**	-.08	.12	.12	.00	.22	-.28	.27	**.72**
Physical/Manual Labor (TH4)	.03	.14	.12	.01	.11	-.09	-.04	**.80**	-.16	-.06	.29	-.10	-.07	-.08	.17	**.71**
Athletics (TH5)	.29	-.04	-.36	.22	-.14	.31	.06	**.16**	.01	-.05	.20	.16	-.07	.23	-.23	**.57**
Protective Service (Th6)	.18	-.03	-.16	.09	.04	.10	.13	**.42**	.05	-.02	-.39	-.07	-.04	.42	-.01	**.56**

Note. *n* = 88. SR scores = Self-reported interest scores; MA scores = Machine-inferred interest scores. S = Health Science; E = Creative Expression; TE = Technology; P = People; O = Organization; I = Influence; N = Nature; TH = Things

Table 4

Latent Factor Correlations of Self-Reported and Machine-Inferred Interest Dimension Scores Derived from the Training and Test Samples

Interest dimensions	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. S (SR)		.13	.22	.24	-.02	.13	.39	.20	**.74**	.05	.10	.16	-.10	-.08	.37	-.08
2. E (SR)	.09		.11	.38	.09	.38	.31	.19	.06	**.63**	.00	.21	.01	.09	.28	-.02
3. TE (SR)	.20	.03		-.06	.22	.27	.24	.53	-.01	-.03	**.78**	-.17	.26	.12	.17	.62
4. P (SR)	.11	.22	-.03		.09	.31	.07	.09	.12	.14	-.14	**.66**	.04	.25	.03	-.15
5. O (SR)	.30	.07	-.06	.10		.41	.05	.11	-.05	-.06	-.06	.01	**.66**	.29	-.22	.10
6. I (SR)	.24	.16	-.04	.15	.38		.02	.28	-.05	.01	.04	.11	.31	**.65**	-.07	.11
7. N (SR)	.29	.27	.13	.05	.03	.00		.24	.23	.42	.13	.02	.08	-.18	**.71**	.12
8. TH (SR)	.41	.28	.25	.02	.34	.16	.36		-.04	-.07	.30	-.09	.04	.10	.25	**.62**
9. S (MA)	**.54**	.02	.26	.24	-.10	-.15	.09	.05		.06	.02	.23	-.12	-.13	.39	-.19
10. E (MA)	-.23	**.21**	-.19	.14	-.22	-.12	-.06	-.18	.04		.01	.29	.10	-.15	.46	-.10
11. TE (MA)	.15	.05	**.26**	-.25	.07	.13	.17	.35	.04	-.04		-.19	.12	-.05	.25	.57
12. P (MA)	.13	.06	.04	**.44**	.14	.15	-.22	-.11	.23	.39	-.17		.11	.22	.07	-.22
13. O (MA)	-.04	-.24	-.21	-.17	**.27**	.09	.05	-.01	-.14	.08	.23	.06		.32	-.12	.21
14. I (MA)	-.01	.00	-.07	.16	.21	**.40**	-.19	.06	-.12	-.02	-.11	.18	.29		-.18	.10
15. N (MA)	.20	.24	.21	.18	-.03	-.19	**.23**	.43	.32	.20	.16	.10	-.20	-.10		.18
16. TH (MA)	.23	-.11	.15	.06	-.06	.12	.34	**.49**	-.05	.03	.39	-.14	.25	.06	.17	

Note. (SR) = self-reported interest scores; (MA) = machine-inferred interest scores. Correlation coefficients above the diagonal are derived from the training sample (n = 229); correlation coefficients below the diagonal are derived from the test sample (n = 88). Bolded are convergent correlation coefficients.

Table 5

Means, Standard Deviations, and Correlations Among Self-Reported and Machine-Inferred Interest Dimension Scores Derived from the Training and Test Samples

Interest dimensions	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	*M*	*SD*
1. S (SR)	2.33	.99		.12	.33**	.35**	.15*	.12	.47**	.22**	**.75****	.07	.10	.27**	-.13	-.14*	.37**	-.02	3.03	1.15
2. E (SR)	2.93	.90	.09		.20**	.52**	.28**	.42**	.42**	.30**	.05	**.64****	.02	.30**	-.02	.09	.33**	-.01	3.05	.95
3. TE (SR)	2.05	.75	.69**	.19		.23**	.44**	.34**	.39**	.65**	.10	.01	**.76****	-.04	.25**	.13*	.22**	.59**	2.69	1.02
4. P (SR)	2.99	.79	.45**	.54**	.40**		.47**	.55**	.46**	.29**	.18**	.28**	-.01	**.65****	.21**	.29**	.24**	-.04	3.11	.83
5. O (SR)	3.04	.88	.37**	.12	.44**	.41**		.72**	.22**	.47**	-.08	-.03	.17**	.19**	**.65****	.49**	-.03	.18**	2.51	.90
6. I (SR)	3.44	.89	.28**	.23*	.28**	.48**	.67**		.19**	.39**	-.09	.06	.10	.27**	.49**	**.65****	-.05	.11	2.76	.91
7. N (SR)	2.50	.93	.45**	.43**	.49**	.49**	.20	.17		.49**	.31**	.36**	.18**	.27**	-.04	-.07	**.69****	.24**	2.93	.95
8. TH (SR)	2.23	.76	.52**	.35**	.67**	.45**	.39**	.33**	.68**		.01	.00	.46**	-.03	.24**	.15*	.26**	**.64****	2.43	.79
9. S (MA)	2.58	.42	**.56****	.07	.32**	.17	-.07	-.14	.18	.17		.11	.05	.35**	-.25**	-.26**	.48**	-.11	3.03	.57
10. E (MA)	3.00	.26	-.20	**.26***	-.31**	.04	-.18	-.12	.02	-.15	-.03		-.01	.42**	-.16*	-.07	.59**	-.08	3.05	.29
11. TE (MA)	2.47	.37	.26*	.00	**.40****	.10	.09	.10	.32**	.43**	.26*	-.03		-.14*	.25**	.07	.24**	.79**	2.69	.52
12. P (MA)	3.09	.21	.18	.11	-.01	**.26***	.17	.13	-.06	-.06	.27*	.51**	-.19		.26**	.37**	.35**	-.24**	3.11	.26
13. O (MA)	2.68	.25	-.01	-.15	.14	.05	**.33****	.32**	-.04	.09	-.27*	-.14	.27*	.16		.79**	-.14*	.29**	2.51	.26
14. I (MA)	2.99	.22	-.09	-.08	.04	.08	.28**	**.36****	-.12	.07	-.29**	.13	.14	.32**	.81**		-.22**	.12	2.76	.27
15. N (MA)	2.67	.31	.28**	.16	.19	.26*	-.10	-.17	**.44****	.34**	.43**	.48**	.28**	.41**	-.15	-.13		.30**	2.93	.37
16. TH (MA)	2.37	.22	.25*	-.07	.42**	.13	.09	.11	.42**	**.57****	-.01	-.12	.72**	-.24*	.27*	.15	.40**		2.43	.25

Note. (SR) = self-reported interest scores; (MA) = machine-inferred interest scores. S = Health Science; E = Creative Expression; TE = Technology; P = People; O = Organization; I = Influence; N = Nature; TH = Things. Correlation coefficients above the diagonal are based on the training sample (n = 229); correlation coefficients below the diagonal are based on the test sample (n = 88). Bolded are convergent correlation coefficients.

Table 6

Convergent and Discriminant Indices for Machine-Inferred Vocational Interest Dimensions Scores

	C1	D1	D2	D2a	MV	MVa
Latent interest dimension scores						
Training sample (*n* = 229)	.68	.55	.49	.50	.06	.05
Test sample (*n* = 88)	.36	.22	.20	.21	.02	.01
Manifest interest dimension scores						
Training sample (*n* = 229)	.68	.51	.36	.41	.15	.10
Test sample (*n* = 88)	.40	.24	.06	.13	.18	.11
Fan et al. (2023; training sample; latent level)	.59	.48	.43	.39	.05	.09
Park et al. (2015; test sample)	.38	.27	.15	.10	.11	.16
Hickman et al. (2022, training sample)	.12	.04	-.05	.01	.09	.04
Marinucci et al. (2018; training sample)	.21		-.03	-.07		

Note. C1(convergence index) = average of absolute monotrait-heteromethod correlations; D1(the first discriminant index) = average of absolute monotrait-heteromethod correlations – average of absolute heterotrait-heteromethod correlations; D2 (the second discriminant index) = average of absolute monotrait-heteromethod correlations – average of absolute heterotrait-monomethod correlations; D2a (discrimination index derived using only machine learning correlations) = average of absolute heterotrait-machine learning method correlations; MV (method variance) = average of absolute heterotrait-monomethod correlations – average of heterotrait-heteromethod correlations; MVa = method variance of machine learning method.

Table 7

Means, Standard Deviations, and Correlations of Two Congruence Indices with Criterion Variables

Criterion variables	*M*	*SD*	1	2	3	4	5	6	7	8	9
1. SR congruence	.67	.30	1.00								
2. MA congruence	.76	.21	.45**	1.00							
3. Major satisfaction	4.43	.85	.47**	.03	1.00						
4. Withdrawal intention	1.31	.85	-.28*	-.14	-.28*	1.00					
5. Absenteeism	1.57	.87	-.23	.07	-.41**	-.01	1.00				
6. Health complaints	26.95	7.55	.21	.30*	.25	-.00	.14	1.00			
7. Subjective health	4.59	1.42	.10	-.09	.02	-.06	-.01	-.32*	1.00		
8. GPA	3.28	.49	.17	.12	.17	-.31*	-.33*	-.06	.12	1.00	
9. High school GPA	3.62	.49	.12	.12	.28*	-.38**	-.08	.10	.16	.61**	1.00

Note. SR Congruence = congruence index calculated using self-reported interest dimension scores; MA Congruence = congruence index calculated uisng machine-inferred interest dimension scores. * indicates $p < .05$. ** indicates $p < .01$.

Table 8

Partial Bivariate Correlations of Two Congruence Indices with Criterion Variables Controlling for High School GPA and P Values

Criterion variables	SR Congruence	*p* value	MA Congruence	*p* value
Major satisfaction	-.02	.91	.12	.40
Withdrawal intention	.01	.91	.07	.59
Absenteeism	.10	.45	.13	.35
Health complaints	.10	.46	.10	.48
Subjective health	.11	.44	.14	.32
GPA	.02	.90	.06	.66

Note. $n = 56$. SR Congruence = congruence index calculated using self-reported interest dimension scores; MA Congruence = congruence index calculated uisng machine-inferred interest dimension scores.

Table 9

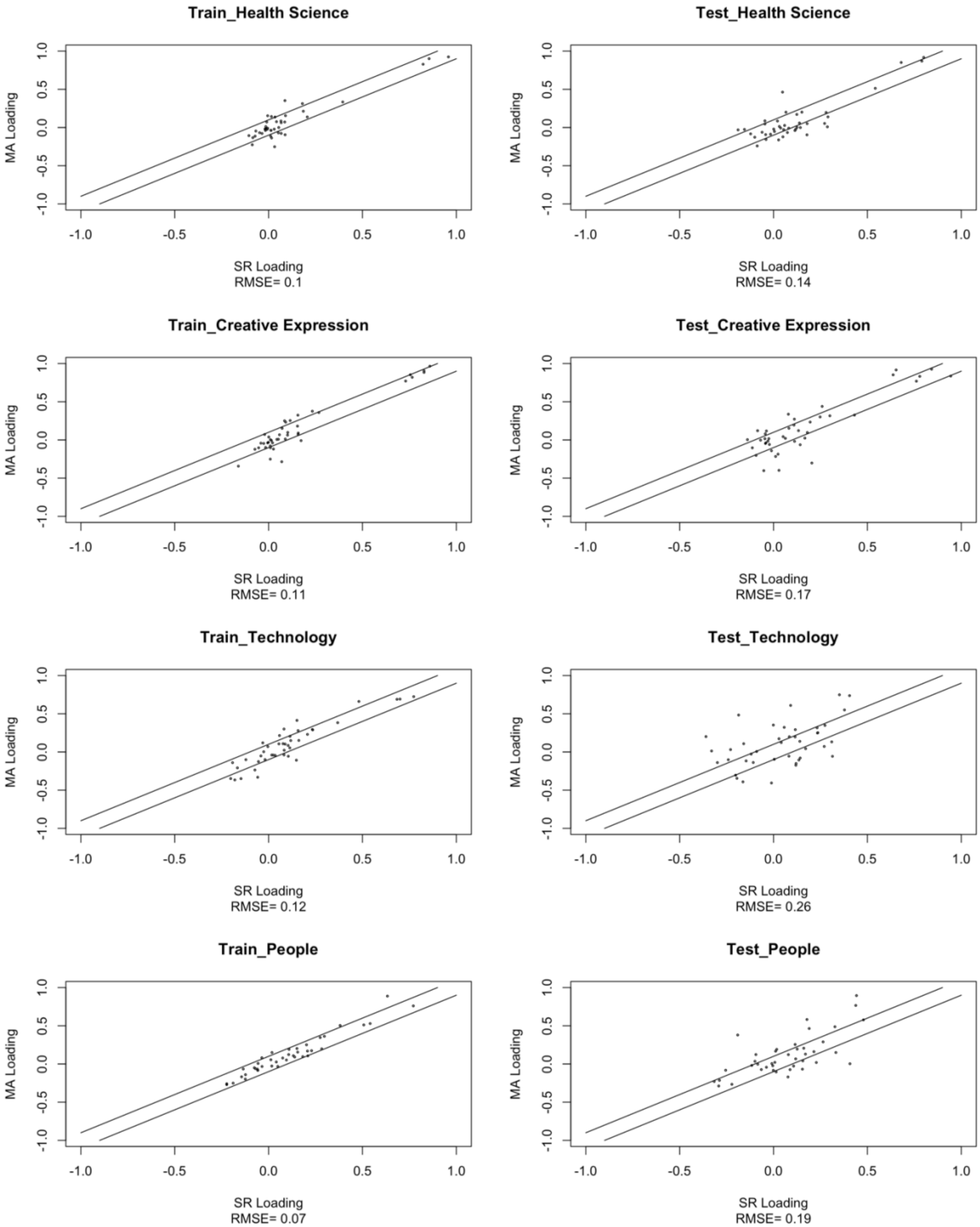
Regression of Major Satisfaction, Withdrawal Intention, Absenteeism, Health Complaints, Subjective Health, and Cumulative GPA on High School GPA, Congruence Indices Derived from Self-Reported Interest Dimension Scores and Machine-Inferred Interest Dimension Scores

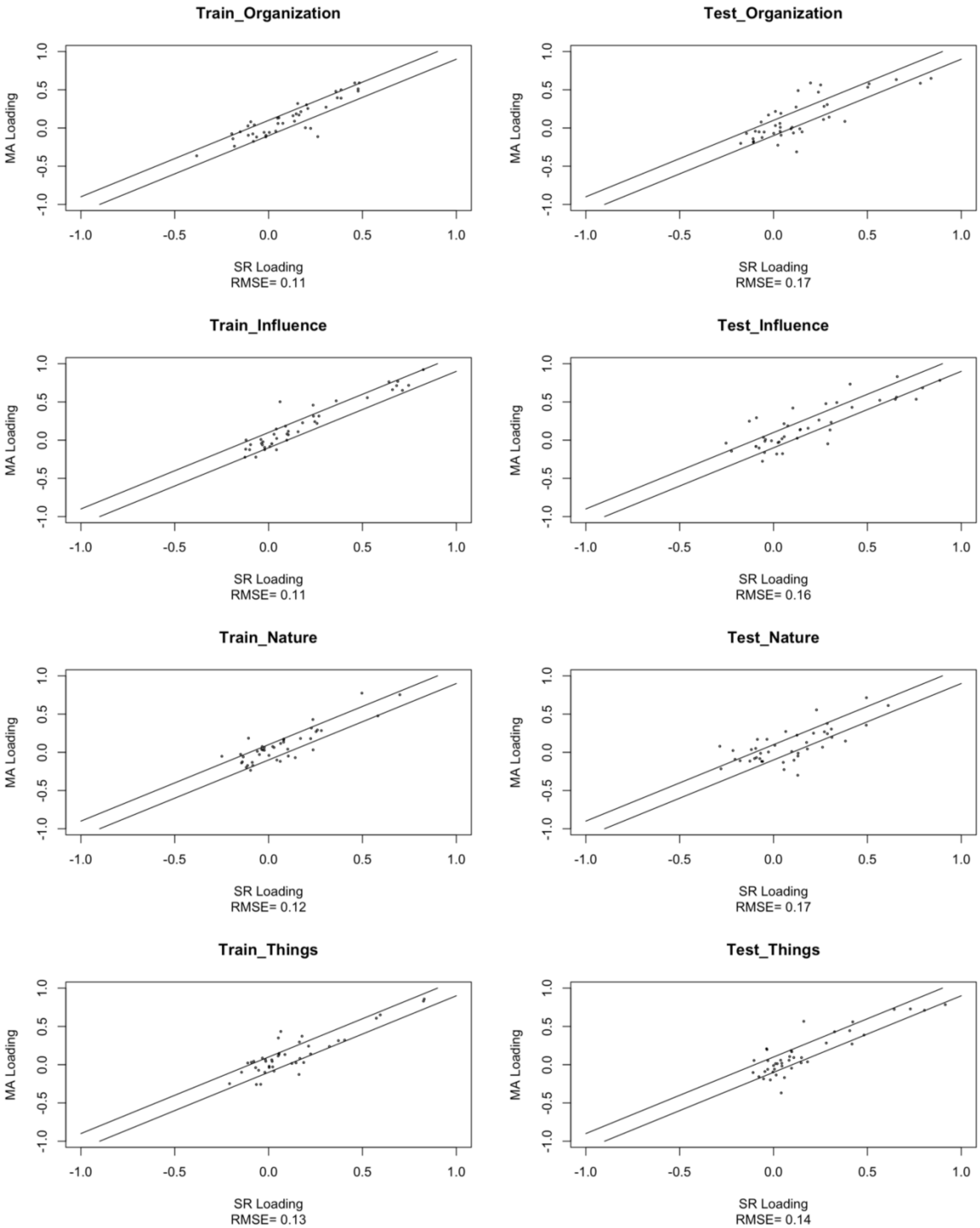
Criterion Variables		High school GPA	SR Congruence	MA Congruence	R^2	ΔR^2
	Step 1	.28*(.13)			.08*	
Major satisfaction	Step 2	.23(.12)	.44**(.12)		.27**	.19**
	Step 3	.24*(.11)	.55**(.13)	-.24(.13)	.32**	.05
	Step 1	-.38**(.12)			.15**	
Withdrawal intention	Step 2	-.35**(.12)	-.24(.13)		.20**	.05
	Step 3	-.35**(.12)	-.24(.14)	.01(.10)	.20**	.00
	Step 1	-.08(.13)			.01	
Absenteeism	Step 2	-.05(.12)	-.22(.13)		.06	.05
	Step 3	-.07(.13)	-.32*(.15)	.22(.14)	.10	.04
	Step 1	.10(.14)			.01	
Health complaints	Step 2	.07(.13)	.20(.14)		.05	.04
	Step 3	.05(.12)	.08(.14)	.25(.15)	.10	.05
	Step 1	.16(.14)			.02	
Subjective health	Step 2	.15(.14)	.08(.13)		.03	.01
	Step 3	.16(.14)	.16(.15)	-.18(.15)	.06	.03
	Step 1	.61**(.11)			.38***	
Cumulative GPA	Step 2	.60**(.11)	.10(.11)		.39***	.01
	Step 3	.60**(.11)	.10(.13)	.00(.00)	.39***	.00

Note. $n = 56$; SR Congruence = congruence index calculated using self-reported interest dimension scores; MA Congruence = congruence index calculated uisng machine-inferred interest dimension scores. * $p < .05$. ** $p < .01$.

Figure 1

Plots of Factor Loadings of Self-Reported and Machine-Inferred Scale Scores and Root Mean Squared Errors Across the Eight Interest Dimensions





Note. RMSE = Root mean squared errors. Train = training sample; Test = test sample. SR = self-reported scale scores; MA = machine inferred scale scores.

Appendix A

AI Chatbot Interview Questions

Warm-up Questions

1. I am from California. Where are you from?
2. Let's expand a bit on that...Where did you grow up? What was the place like?

Topic 1: Activities you enjoy

3. What do you love to do outside of studying for your classes?
4. Why do you enjoy doing this(these)?

Topic 2: Your favorite location

5. Just like a person, every place has its own unique personality. Where do you want to live in the future, a big city or a small town? Could you give one example of such a place?
6. Why would you want to live there?
7. What would you like to do in your favorite location?

Topic 3: People/things

8. There are different types of work activities and people may have their own preferences for those activities. For example, some people like having discussions with others to help them solve their problems, while other people may prefer working alongside machines to repair/build/operate something. Could you talk about what is your preference for either of these two types? You can type "Neither of these interests me" if you like neither of them.
9. [Only ask this question if the participant indicates "neither" for Question 8] If neither of these activities looks appealing, what kind of work activities are more interesting to you?
10. Could you provide mean example or examples of such anactivity that interest(s) you?
11. Why does this activity appeal to you?

Topic 4: Career goal

12. What is your long-term career goal?
13. Why did you choose this goal? Put differently, what aspects of this career goal��are most appealing to you?
14. What actions will you take to achieve your long-term goal?

Topic 5: Creative expression/Technology

15. As we just talked about, there are different types of work activities and people may have their own preferences for those activities. Could you talk about what is your preference for either of the following types? Creating pieces of art that are pleasing or could express your idea(s) OR developing or using technology to improve the way things can be done. You can type "Neither of these interests me" if you like neither of them.
16. [Only ask this question if the participant indicates "neither" for Question 15] If neither of these activities looks appealing, what kind of activities are more interesting to you?

17. Could you provide me an example or examples of such an activity that interest(s) you?
18. Why does this activity appeal to you?

Topic 6: what is your passion?

19. Passion is contagious. What are you passionate about?
20. How did this passion start?
21. What have you been doing to pursue your passion?

Topic 7: Health science/Organization

22. Again, let's talk about different types of work activities and your preference. Could you talk about what is your preference for either of the following types? Conducting research on health-related issues or providing medical services to help cure people OR planning and organizing things/people to facilitate organization activities. You can type "Neither of these interests me" if you like neither of them.
23. [Only ask this question if the participant indicates “neither” for Question 22] If neither of these activities looks appealing, what kind of activities are more interesting to you?
24. Could you provide me an example or examples of such an activity that interest(s) you?
25. Why does this activity appeal to you?

Topic 8: Favorite course in college

26. What's your favorite course in college?
27. What aspects of the course appeal to you?
28. How did you do in that course?

Topic 9: Influence/Nature

29. Again, let's talk about different types of work activities and your preference. Could you talk about what is your preference for either of the following types? Leading, persuading or influencing people to achieve the goal of your group OR doing outdoor work related to nature or agriculture. You can type "Neither of these interests me" if you don't like either.
30. [Only ask this question if the participant indicates “neither” for Question 29] If neither of these activities look appealing, what kind of activities are more interesting to you?
31. Could you provide me an example or examples of such an activity that interest(s) you?
32. Why does this activity appeal to you?

Topic 10: Your college major

33. Have you picked your major yet?
34. [Only ask this question if the participant indicates “yes” for Question 33] What is it? Why do you choose this major?
35. [Only ask this question if the participant indicates “no” for Question 33] When must you pick your major?
36. [Only ask this question if the participant has answered Question 35] What are some possible majors you are

thinking about?

37. [Only ask this question if the participant has listed several majors in their answer for Question 35] What aspects of these majors attract you?
38. [Only ask this question if the participant did not list any major in their answer for Question 35] Why is it difficult for you to make up your mind?

Topic 11: Values

39. Do you think that a job/working is the most important thing in your life?
40. [Only ask this question if the participant indicates “yes” for Question 39] Why do you value your job/work the most?
41. [Only ask this question if the participant indicates “yes” for Question 40] What do you value most in your life? Why?
42. Do you have any personal beliefs about life? In other words, what are your beliefs about yourself, others, the world, or the future?
43. I would like to know more about things you think are important in your life. Which of the following is more important to you? Or both? If neither of these is important to you, please type "Neither of these is important to me". Protecting and enhancing the welfare of people and nature OR achieving personal success and/or dominating people and resources.
44. [Only ask this question if the participant indicates “neither” for Question 43] If neither of these is important to you, what is important to you?
45. Why is it important or why are they important?
46. Could you tell me a little bit more about things that are important to you? Which of the following is more important to you? Or both? If neither of these is important to you, please type "Neither of these is important to me". Maintaining traditions and relationships, and/or prioritizing safety and stability of yourself and the society OR seeking excitement in life, determining your own actions, and cultivating your own ideas and abilities independently.
47. [Only ask this question if the participant indicates “neither” for Question 46] If neither of these is important to you, what is important to you?
48. Why is it important or why are they important?

Topic 12: Work settings

49. What would give you the greatest satisfaction at work/school?
50. Under what kinds of conditions do you work best? (e.g., formal vs. casual, team-based vs. individual…)
51. Recall a challenge you have faced at work or at school and how you handled it.
52. Could you tell me a positive trait you have?
53. Could you tell me an example of a time you express that trait?

54. Could you tell me a negative trait you have?

55. Could you tell me an example of a time you express that trait?

Appendix B

Table of Rotated Factor Loading Matrices Using Set-Exploratory Structural Equation Model Derived from the Training Sample

Table A3

Rotated Factor Loading Matrices Using Set-Exploratory Structural Equation Model Derived from the Training Sample

	SR scores								MA scores							
Basic interest scales	S	E	TE	P	O	I	N	TH	S	E	TE	P	O	I	N	TH
Life Science (S1)	**.82**	.02	.24	-.14	-.11	.02	.08	-.02	**.83**	-.01	.29	-.07	-.10	-.05	.17	.04
Medical Science (S2)	**.96**	.00	.06	-.06	.00	-.04	-.02	-.07	**.92**	-.03	.11	.01	-.06	-.04	.06	-.04
Health Care Service (S3)	**.86**	-.02	-.17	.21	.22	-.10	-.05	.00	**.90**	-.10	-.21	.17	-.01	-.06	-.03	-.04
Media (E1)	-.09	**.83**	-.02	.01	-.08	.18	-.15	.02	-.23	**.90**	-.10	.06	-.08	.23	-.03	-.03
Applied Arits & Design (E2)	-.01	**.76**	.09	-.12	.14	-.03	.22	.02	-.02	**.85**	.10	-.14	.09	-.02	.18	.04
Music (E3)	.09	**.86**	.16	-.04	-.02	-.12	-.14	.06	.07	**.97**	.28	.08	-.12	.00	-.14	.15
Visual Arts (E4)	.03	**.73**	-.04	-.12	-.02	-.12	.17	-.05	.14	**.77**	-.05	-.20	-.06	-.12	.18	-.07
Performing Arts (E5)	-.02	**.83**	-.15	.02	-.06	.10	-.14	.06	-.04	**.88**	-.35	.15	-.12	.08	-.06	.13
Creative Writing (E6)	-.05	**.77**	-.12	.11	-.08	.04	-.04	-.02	-.07	**.82**	-.10	.19	-.18	.15	.04	-.10
Culinary Art (E7)	.07	**.27**	.15	.15	.16	.04	.26	-.04	.06	**.36**	.41	.20	.32	-.13	.29	-.26
Engineering (TE1)	.01	.04	**.68**	.04	.05	.11	.10	.18	-.04	.01	**.69**	.03	.06	.12	.04	.37
Physical Science (TE2)	.40	.06	**.48**	-.04	-.20	.11	.23	.05	.33	.07	**.66**	-.03	-.08	.07	.32	.11
Information Technology (TE3)	-.07	-.01	**.77**	.05	.08	-.04	-.14	.16	-.11	-.04	**.72**	-.03	.06	.01	-.13	.30
Mathematics/Statistics (TE4)	.04	-.04	**.70**	.08	.05	-.02	-.12	.05	.08	-.04	**.69**	.08	.13	-.10	-.20	.35
Social Science (P1)	.19	.09	.00	**.54**	-.11	.26	.25	-.07	.21	.23	.07	**.53**	.03	.22	.27	-.26
Humanities & Foreign Language (P2)	.07	.23	.06	**.30**	-.19	.25	.24	.12	.09	.38	.22	**.36**	-.14	.24	.43	.02
Teaching/Education (P3)	-.11	.16	.11	**.63**	.17	.02	.04	-.03	-.11	.09	.05	**.89**	.21	-.05	-.10	.09
Social Service (P4)	.09	-.02	.02	**.77**	.20	.01	.24	-.21	.16	.07	-.04	**.76**	.30	-.08	.03	-.25
Religious Activities (P5)	.18	.07	-.07	**.28**	.26	.06	-.25	.15	.31	-.29	-.24	**.20**	-.11	.50	-.05	.02
Human Resources (O1)	.01	.01	-.05	.20	**.48**	.36	-.06	.17	-.04	-.10	-.13	.25	**.48**	.51	.01	.08
Personal Service (O2)	.07	.16	-.06	.38	**.46**	.16	.02	.03	-.08	.07	-.33	.50	**.59**	.11	.08	-.08
Accounting (O3)	.03	-.16	.16	-.06	**.48**	.27	-.03	.22	-.02	-.34	.15	-.08	**.51**	.32	.03	.14
Office Work (O4)	.03	.09	.08	.11	**.48**	.10	-.03	.32	-.25	.01	.30	.12	**.59**	.00	.08	.24

Finance (O5)	.05	-.07	.23	-.14	**.39**	.53	.00	.09	.00	-.12	.29	-.17	**.39**	.56	-.04	.13
Management/Administration (I1)	-.08	-.05	.11	.13	.21	**.66**	-.11	.02	-.13	-.10	.21	.11	.26	**.66**	-.17	.06
Business Initiatives (I2)	-.04	.01	.08	-.22	.39	**.71**	.11	-.08	-.08	.00	.02	-.27	.50	**.65**	-.05	.04
Marketing/Advertising (I3)	-.02	.16	.04	-.19	.37	**.68**	.06	-.09	-.02	.32	-.05	-.25	.40	**.71**	-.12	.03
Professional Advising (I4)	.02	.10	.11	.51	.31	**.24**	-.08	-.02	.14	.06	-.06	.51	.27	**.46**	-.17	.07
Public Speaking (I5)	-.01	.15	.12	.28	-.15	**.64**	-.09	-.14	.07	.18	.15	.35	-.05	**.76**	-.13	-.10
Sales (I6)	-.02	.07	.03	-.22	.36	**.75**	.14	-.11	-.02	.15	-.04	-.26	.48	**.72**	-.07	.02
Politics (I7)	.01	.01	-.03	.18	-.38	**.82**	-.11	.19	-.11	-.25	.00	.10	-.36	**.92**	.19	.03
Law (I8)	.09	.01	-.19	.16	-.18	**.69**	-.02	.16	-.10	-.08	-.14	.16	-.24	**.77**	.03	-.13
Agriculture (N1)	.21	-.03	.10	.10	-.08	.04	**.50**	.41	.14	-.05	.09	.05	.04	.02	**.77**	.32
Outdoors (N2)	-.01	.09	.08	.21	-.10	-.02	**.70**	.21	-.01	.25	-.04	.11	.08	-.12	**.75**	.24
Animal Service (N3)	.09	.12	-.18	.23	.06	-.07	**.58**	.00	.35	.25	-.37	.17	.13	-.22	**.48**	-.02
Mechanics/Electronics (TH1)	-.02	.02	.37	-.06	-.01	.03	.06	**.60**	.01	-.04	.38	-.07	-.10	.08	.12	**.65**
Transportation/Machine Operation (TH2)	-.02	.05	.08	-.08	.04	-.03	.08	**.83**	-.02	.01	.11	-.05	-.04	-.06	.16	**.86**
Construction/Woodwork (TH3)	-.07	.10	.21	-.07	.15	-.13	.28	**.57**	-.05	.10	.23	-.06	.19	-.22	.28	**.61**
Physical/Manual Labor (TH4)	.05	.00	-.03	-.06	.20	-.10	.08	**.83**	-.07	.04	.12	-.08	.00	-.13	.13	**.83**
Athletics (TH5)	-.01	.17	.15	.14	.16	.09	-.10	**.07**	.15	-.01	-.11	.09	.17	.18	-.24	**.43**
Protective Service (Th6)	.02	.03	-.20	.02	.13	.24	.03	**.37**	-.14	-.12	-.35	-.03	.16	.31	.06	**.31**

Note. *n* = 229. SR scores = Self-reported interest scores; MA scores = Machine-inferred interest scores. S = Health Science; E = Creative Expression; TE = Technology; P = People; O = Organization; I = Influence; N = Nature; TH = Things