Special Collection: Papers From the 16th European Conference on Psychological Assessment (ECPA 2025)

Psychometric Evaluation of the AI-Based Career Matching Platform meinBeruf.ch

Evidence for Reliability, Construct Validity, and Gender Fairness

Martin Steppan[1], George Jogho[2], Isabelle Ruchti[3], and Benedikt Hell[4]

[1]Computational Psychology Lab, Basel, Switzerland
[2]Katholische Hochschule Freiburg, Freiburg im Breisgau, Germany
[3]Faculty of Psychology, University of Basel, Basel, Switzerland
[4]School of Applied Psychology, University of Applied Sciences and Arts Northwestern Switzerland FHNW, Olten, Switzerland

Abstract: *Objective:* This study examined the preliminary psychometric quality and construct validity of the meinBeruf.ch assessment system and evaluated gender-related fairness at the level of input variables and algorithmic recommendations. *Methods:* The sample comprised 914 users of the meinBeruf.ch platform. Participants completed standardized personality (BFI-2), interest (RIASEC), cognitive (HMT), and domain-specific skill measures, followed by algorithmic career recommendations. Reliability and construct validity were assessed using internal consistency coefficients and correlations between psychological scales and industry-fit scores. Gender fairness was examined by testing whether self-reported gender added incremental predictive value beyond psychological profiles, and by reporting supplementary input-variable analyses of gender differences, DIF, measurement invariance, and a technical counterfactual audit of the ranking layer. *Results:* The psychological scales demonstrated adequate to high internal consistencies, and correlations with algorithmic recommendations showed theoretically consistent patterns. Adding gender to random-forest prediction models changed predictive accuracy only minimally (M $|\Delta R^2|$ = .011), indicating that gender provided little systematic predictive value beyond the psychological profile. Supplementary analyses documented expected gender differences in some input variables and no counterfactual rank changes when only the gender label was varied. *Limitations:* The study used a self-selected online sample, relied primarily on cross-sectional platform data, and lacked long-term criterion outcomes such as later career choice, satisfaction, or persistence. *Conclusion:* The meinBeruf.ch assessment system provides preliminary evidence for psychometric quality and construct validity. The algorithmic ranking layer behaved technically gender-fair, while gender-related differences in personality, interests, and self-estimated skills remain important for interpretation and future research, especially in STEM-related vocational guidance.

Keywords: career counseling, AI-based career matching, psychometrics, construct validity, algorithmic fairness

Vocational guidance is a multifactorial task (Savickas, 2013). Digital systems can support this task by organizing large bodies of psychological and occupational information and by making fit hypotheses explicit. This does not imply that such systems replace human counselors or capture the full complexity of career development. Rather, they can provide structured information that may complement counseling, user agency, and lived career experience. Because self-descriptions can be biased or incomplete (Neubauer & Hofer, 2021), standardized psychological assessment provides one way to generate comparable input for algorithmic recommendations, while still remaining subject to the general limitations of self-report.

A variety of established vocational guidance instruments already exist, including EXPLORIX – the Swiss adaptation of Holland's RIASEC model – as well as other electronic interest inventories and matching systems based on person–environment fit theory (Hell & Buri, 2023). While these systems are widely used and

https://doi.org/10.1027/2698-1866/a000147

effective, they are based on fixed scoring rules and lack the adaptive flexibility of modern AI approaches. Furthermore, person–environment fit models have conceptual limits, since real career decisions are influenced by individual biographies, contextual opportunities and constraints, and the process of active career construction (Savickas, 2013).

Recent machine-learning systems may help combine multiple psychological and occupational data sources and provide individualized profile-congruence estimates (José-García et al., 2023; Manganello et al., 2025). In addition, large language models can process free-text descriptions of personality, values, and motivations, adding nuance beyond fixed questionnaire formats (Rosenfelder et al., 2025).

At the same time, concerns about algorithmic bias, including gender, racial, ethnic, and socioeconomic bias, are well-documented in educational and occupational contexts (Barocas et al., 2023; Binns, 2018; Menis–Mastromichalakis et al., 2025). Gender fairness in vocational guidance requires a distinction between unfair measurement or algorithmic bias and meaningful group differences in vocational interests and career preferences, which are well-documented in interest research (Bergmann, 2003; Gottfredson, 1982; Su et al., 2009; Wetzel & Hell, 2013). The goal of the present study was therefore twofold: (1) to examine the psychometric quality and construct validity of the meinBeruf.ch assessment system, and (2) to evaluate gender-related fairness at the input-variable and algorithmic recommendation levels. Together, these analyses provide an initial evaluation of measurement quality and procedural fairness of the platform.

Methods

Participants and Procedure

Participants were recruited through an evaluation study conducted at the University of Basel (Switzerland) and the Macromedia University of Applied Sciences in Freiburg im Breisgau (Germany), as well as through individuals who independently discovered the meinBeruf.ch website and completed the assessment voluntarily. In total, 914 users took part, of whom 912 provided at least one usable scale score. Because participants did not complete every module, analytic sample sizes varied by measure (see Table 1). The M_{age} was 26.8 years (SD = 10.7). Among participants with available gender information (n = 821), 63.2% identified as female, 35.2% as male, and 1.6% as diverse or third gender.

Measures

The assessment battery consisted of standardized personality, interest, cognitive, and domain-specific skill measures

Table 1. Internal consistency indices for scales used in meinBeruf.ch (Reliability)

Scale	Instrument	k	N	α	ωt / SB	Reference
Extraversion	BFI-2	12	824	.87	.87	(.83)
Agreeableness	BFI-2	12	824	.81	.81	(.81)
Conscientiousness	BFI-2	12	824	.86	.86	(.84)
Neuroticism	BFI-2	12	823	.87	.87	(.85)
Openness	BFI-2	12	823	.86	.86	(.87)
Realistic	RIASEC	8	442	.83	.84	(.88–.90)
Investigative	RIASEC	8	442	.88	.88	(.90–.91)
Social	RIASEC	8	442	.82	.83	(.84–.89)
Enterprising	RIASEC	8	442	.76	.76	(.84–.88)
Artistic	RIASEC	8	442	.84	.84	(.89)
Conventional	RIASEC	8	442	.86	.86	(.91–.93)
Fluid Intelligence	HMT	20[a]	480	.79	.80	(.78)
Resilience	Domain-specific brief scales	3	511	.65	.66	—
Math and Digital Skills	Domain-specific brief scales	2	511	.46	SB = .46	—
Social and Environmental Engagement	Domain-specific brief scales	2	510	.71	SB = .71	—
Social Skills	Domain-specific brief scales	4	511	.58	.62	—
Language Skills	Domain-specific brief scales	3	511	.56	.57	—

Note. k = number of items; N = number of participants; α = Cronbach's alpha; ωt = McDonald's omega total; Reference = reliability reported by test authors. For two-item scales, McDonald's ωt is not estimable; therefore, the Spearman–Brown coefficient is reported in the ωt / SB column. [a]The HMT scale comprises 20 scored items; due to a technical issue, the 20th item was not shown to all participants in the exported data used for these supplementary checks, but this issue has since been fixed.

administered digitally as part of the meinBeruf.ch platform. Personality traits were assessed with the German version of the Big Five Inventory-2 (Danner et al., 2016). Vocational interests were captured using the public-domain RIASEC inventory from OpenPsychometrics.org, based on Holland's (1997) six-dimensional interest model (Liao et al., 2008). Fluid reasoning ability was assessed with the Hagen Matrices Test, a matrix reasoning task measuring inductive and abstract problem solving (Heydasch, 2014). In addition, the platform includes several short domain-specific scales developed for meinBeruf.ch: Resilience, Mathematical and Digital Skills, Social and Environmental Engagement, Social Skills, and Language Skills. These brief scales are not intended as precise stand-alone diagnostic instruments; instead, they provide compact profile indicators that add nuance to multivariate sorting and recommendation. To avoid introducing gender-based stratification, scale scores were normed using gender-unspecific norms, and gender was not used in the standardization process.

AI-Based Recommendation System meinBeruf.ch

The vocational guidance platform meinBeruf.ch combines standardized psychological tests with an AI-supported recommendation system called the JOBMATCH algorithm (Steppan, 2025). After completing the assessment, users receive personalized fits for more than 1,000 professions, numerous study fields, 39 occupational branches from the SWISSDOC database (Schweizerisches Dienstleistungszentrum Berufsbildung [SDBB], 2025), and international recommendations based on O*NET (Hilton & Tippins, 2010). The algorithm operates anonymously and does not use demographic information such as age or gender. It is best understood as a profile-congruence and ranking procedure that combines distance-based similarity measures between user and occupational profiles with semantic matching of free-text information. A detailed technical description of the JOBMATCH algorithm is available on the meinBeruf.ch website.

Statistical Analyses

Reliability and Construct Validity

All statistical analyses were conducted in *R* (R Core Team, 2025). Construct validity was examined using Pearson correlations between the psychological scales and the algorithm's recommendation scores, with pairwise complete observations. Reliability was estimated using Cronbach's α and McDonald's ωt for scales with three or more items (McDonald, 2013). For two-item scales, McDonald's ωt is not estimable; therefore, the Spearman–Brown coefficient is reported.

Gender Fairness

To examine gender fairness, we distinguished input-variable fairness from technical fairness of the algorithmic ranking layer. Input-variable analyses included scale-level gender differences, multigroup CFA measurement invariance, and regression-based differential item functioning (DIF) screening. These analyses are reported in Section A of Electronic Supplementary Material 1 (ESM 1). At the algorithmic-output level, we examined whether gender contributes to the algorithm's industry recommendations beyond participants' psychological profiles using random forests (Liaw & Wiener, 2002). The JOBMATCH algorithm itself does not use gender as an input; however, gender could still be associated with recommendations indirectly through gender-differentiated psychological profiles. Random forests provide a flexible test because they can model nonlinear relations and high-order interactions (Breiman, 2001).

We used the 17 psychological scales reported in Table 2 (Big Five domains, RIASEC interests, Fluid Intelligence, and domain-specific brief scales) as predictors. Participants identifying as male or female were included, and gender was coded as a categorical variable.

For each of the 39 industry branches, we estimated two predictive models:

1. Trait-only model: $\hat{y} = f(X)$
2. Trait + gender model: $\hat{y} = f(X, \text{Gender})$

where f(·) denotes a random forest regression function and X represents the vector of psychological traits. Model performance was evaluated using R^2 on a held-out 30% test set. The incremental predictive value of gender was quantified as:

$$\Delta R^2 = R^2(\text{traits} + \text{gender}) - R^2(\text{traits})$$

Values of ΔR^2 close to zero indicate that gender does not improve prediction once psychological traits are accounted for, thereby supporting procedural gender fairness at the level of branch recommendations. In addition, a technical counterfactual audit of the ranking layer was conducted by holding real user profiles constant and varying only the gender label; this audit is reported in Section B of ESM 1.

Results

The results are presented in three parts: Table 1 summarizes the reliability of all scales, Table 2 reports construct validity coefficients with the meinBeruf.ch matching outcomes, and

Table 2. Correlations between the algorithm's industry recommendations and participant measures (construct validity)

Scale	STEM	Humanities	Arts and fashion	Business and trade	Hospitality and beauty	Production	Agriculture
Extraversion	−0.18**	0.32***	0.19**	0.23***	0.35***	−0.32***	−0.05
Agreeableness	−0.16*	0.34***	0.22***	0.03	0.29***	−0.26***	−0.03
Conscientiousness	−0.11	0.05	−0.17**	0.12	0.17**	−0.13*	0.10
Neuroticism	−0.08	0.04	0.13*	−0.29***	0.08	0.13*	0.03
Openness	0.11	0.47***	0.50***	−0.07	0.10	−0.56***	−0.02
Realistic	0.12	−0.28***	−0.19**	−0.10	−0.28***	0.13*	0.12
Investigative	0.28***	0.26***	−0.13*	−0.16*	−0.15*	−0.26***	0.27***
Artistic	−0.03	0.41***	0.52***	−0.22***	0.24***	−0.34***	0.09
Social	−0.30***	0.49***	0.32***	−0.14*	0.49***	−0.31***	0.21***
Enterprising	−0.28***	−0.03	0.20**	0.15*	0.26***	0.12	−0.08
Conventional	−0.15*	−0.34***	−0.16*	−0.03	−0.06	0.30***	−0.03
Fluid Intelligence	0.64***	0.18**	−0.03	0.34***	−0.37***	−0.29***	−0.42***
Resilience	0.14*	0.03	−0.18**	0.03	−0.06	−0.18**	0.20**
Math and Digital Skills	0.38***	−0.12	−0.24***	0.21***	−0.34***	−0.08	−0.20**
Social and Environmental Engagement	−0.07	0.43***	0.07	−0.22***	0.22***	−0.22***	0.26***
Social Skills	−0.19**	0.50***	0.26***	0.25***	0.38***	−0.46***	−0.07
Language Skills	−0.12	0.49***	0.28***	0.27***	0.36***	−0.50***	−0.07

Note. Correlations quantify the extent to which algorithmic industry recommendations align with established psychological constructs, providing evidence of construct validity. Boldface values indicate statistically significant correlations. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

the gender-fairness section summarizes algorithmic-output analyses. Detailed input-variable gender differences, DIF, measurement invariance, and technical counterfactual audit results are reported in ESM 1.

Reliability

Table 1 summarizes internal consistencies and item counts of the applied scales. The Big Five domains showed adequate to high reliability (ω = .81–.87), and the RIASEC interest scales demonstrated similar internal consistency (ω = .76–.88). The fluid reasoning task (HMT) showed acceptable reliability (ω = .80), comparable to previously reported values for online matrix reasoning tests. The very brief domain-specific scales showed more variable reliability (α/Spearman–Brown = .46–.71; ω = .57–.66 for scales with at least three items), consistent with their intended use as compact profile indicators rather than standalone diagnostic instruments.

Construct Validity

Table 2 displays coherent and theoretically meaningful associations between participants' psychological profiles and the algorithm's industry-fit scores. Humanities recommendations were most strongly related to Social Skills (r = .50), Social Interests (r = .49), Language Skills (r = .49), Openness (r = .47), Social and Environmental Engagement (r = .43), and Artistic Interests (r = .41). Arts and Fashion showed the strongest associations with Artistic Interests (r = .52), Openness (r = .50), Social Interests (r = .32), Language Skills (r = .28), and Social Skills (r = .26). Business and Trade recommendations were associated most strongly with Fluid Intelligence (r = .34), Language Skills (r = .27), Social Skills (r = .25), Extraversion (r = .23), and Math and Digital Skills (r = .21). STEM recommendations showed the expected pattern of positive associations with Fluid Intelligence (r = .64), Math and Digital Skills (r = .38), and Investigative Interests (r = .28), and a small positive association with Realistic Interests (r = .12). Production recommendations were negatively associated with Openness (r = −.56), Language Skills (r = −.50), Social Skills (r = −.46), Artistic Interests (r = −.34), Extraversion (r = −.32), and Social Interests (r = −.31), while showing a positive association with Conventional Interests (r = .30). Hospitality and Beauty showed pronounced associations with Social Interests (r = .49), Social Skills (r = .38), Language Skills (r = .36), Extraversion (r = .35), Agreeableness (r = .29), Enterprising Interests (r = .26), and Artistic Interests (r = .24). Agriculture was positively associated with Investigative Interests (r = .27), Social and Environmental Engagement (r = .26), Social Interests (r = .21), and

Resilience (r = .20), and negatively associated with Fluid Intelligence (r = −.42) and Math and Digital Skills (r = −.20).

Overall, the correlations show that the algorithm's recommendations align closely with established personality–interest–environment relations (Liu et al., 2025), providing strong evidence for construct validity.

Gender Fairness

Across the 39 industry branches, the random forest models that included only psychological trait predictors and those that additionally included gender produced highly similar levels of predictive accuracy. The average absolute difference in out-of-sample R^2 between the two models was small (M $|\Delta R^2|$ = .011), indicating that the inclusion of gender changed R^2 only minimally across branches. This pattern suggests that self-reported gender did not add systematic predictive information beyond participants' psychological profiles.

The largest incremental value was observed for the fashion branch (ΔR^2 = .051). Smaller gender-related increments were observed for a limited number of additional branches, whereas most branches showed changes close to zero. The content of ESM 1 expands these findings in two ways. Section A documents expected gender differences in several input variables, including higher Realistic Interests and Math and Digital Skills among men and higher Neuroticism, Agreeableness, and Social and Environmental Engagement among women. Section B reports a real-user counterfactual audit of the ranking layer; changing only the gender label produced no top-10 differences, no top-50 differences, and no rank shifts.

Discussion

The results of this evaluation study provide preliminary evidence that the meinBeruf.ch assessment system has acceptable psychometric properties and produces theoretically coherent career recommendations across a wide range of occupational branches. Reliability indices were adequate for most established scales, construct validity was supported by meaningful correlations between psychological traits and industry-fit scores, and the fairness analyses indicated that the audited ranking layer behaves in a technically gender-fair manner when gender is varied while the psychological profile is held constant.

Scale reliabilities were largely consistent with, and in some cases close to, those reported by the original test developers, indicating that the established scales retained adequate psychometric quality in the study sample and under online administration (Armstrong et al., 2008; Danner et al., 2019; Heydasch, 2014). The Big Five and RIASEC scales demonstrated solid internal consistency. By contrast, the very brief domain-specific measures showed lower and more variable reliability. These short scales should therefore not be interpreted as precise standalone diagnostic instruments. Their intended role is narrower: they provide compact profile features that add nuance to multivariate sorting, for example, whether a person reports stronger digital, language-related, social, or environmental orientations. Future versions of the platform should continue to evaluate and, where useful, expand these short scales.

Construct validity was also supported, with correlations between psychological traits and industry recommendations aligning well with established findings in vocational psychology. The observed patterns correspond closely to Holland's (1997) RIASEC theory, which predicts systematic associations between personality characteristics and preferred work environments. For example, Social and Artistic Interests were most strongly related to creative and interpersonal branches, Realistic Interests to technical and manual fields, and Conscientiousness and Math and Digital Skills to business and technology-oriented domains – patterns consistent with research on person–environment fit and personality–interest relations (Anni et al., 2024; Batista & Gondim, 2022; Etzel et al., 2025; Larson et al., 2002). These findings indicate that the JOBMATCH Algorithm uses psychologically meaningful information and produces recommendations that reflect established vocational choice processes.

The fairness analyses should be interpreted as evidence for procedural and technical gender fairness, not as evidence that all input variables are fully measurement-invariant or that group-level outcomes must be identical. The algorithm does not use gender as an input variable, no gender-specific weights or thresholds are applied, and the same scoring function operates for all users. Consistent with this design, the random forest analyses showed that adding gender did not meaningfully improve prediction once psychological profiles were included, and the counterfactual audit showed unchanged ranks when only the gender label was varied. At the same time, the content of ESM 1 shows substantial gender differences in some interests and self-estimated skills. Such group-level differences are well-documented in vocational psychology, especially for Realistic and Social Interests (Su et al., 2009), and should not be interpreted as algorithmic bias. Rather, they point to an important use case for the tool: making gendered interest profiles visible and supporting research and guidance interventions in areas such as STEM/MINT pathways.

Limitations

Despite these strengths, the study has several limitations. The sample was a convenience sample of volunteers, and not all participants completed the full battery. Missing data were handled by using all available observations for each analysis, but incomplete participation reduces precision and could bias results if missingness is systematic. The evaluation took place in a fully online and unsupervised setting, which may have influenced engagement, particularly on cognitively demanding tasks. The industry branches examined are relatively broad, and results may differ for more fine-grained occupational classifications. Demographic diversity beyond binary gender could not be examined due to the small number of participants identifying as diverse or third gender (n = 13). Finally, because criterion outcomes such as actual career choice, job satisfaction, or educational persistence were not available, the present evaluation cannot establish long-term predictive or criterion validity.

Future Directions

Looking ahead, a key priority is to examine criterion validity by linking algorithmic recommendations to real-world outcomes. In particular, obtaining data from individuals who are highly satisfied or dissatisfied with their careers would allow the system to learn which psychological profiles thrive in which environments – extending beyond fit scores to outcomes such as happiness, financial satisfaction, or meaning in life. We are planning a large-scale study in Switzerland focusing on individuals' career trajectories and satisfaction outcomes, which will allow the algorithm to be trained on a broader set of criterion variables. Beyond person–environment fit, future versions of the platform will also implement the Career Construction Interview and additional psychometric measures. These extensions are intended to make the tool less of a stand-alone decision engine and more of a structured basis for a well-informed conversation between career counselors and clients. Future work should also examine generalizability beyond Switzerland and the D/A/CH context, including international samples and educational systems with different opportunity structures.

Conclusion

Digital career-counseling tools are increasingly complementing traditional vocational guidance (José-García et al., 2023; Manganello et al., 2025; Möttus & Möttus, 2025). Given this trend, the scientific community should evaluate such systems carefully, transparently, and with attention to both psychometric quality and fairness. The present study is an initial step in that direction. It suggests that AI-supported matching can be evaluated with established psychological methods, while also showing that stronger user-centered, criterion-validity, and cross-cultural evidence is needed before broader claims about practical benefit can be made.

Electronic Supplementary Material

The following electronic supplementary material is available at https://doi.org/10.1027/2698-1866/a000147
ESM 1. Gender-Fairness in Input Variables, Technical Gender-Fairness of the Algorithm, and Trait-Block Contributions to Fit Predictions.

References

Anni, K., Vainik, U., & Mõttus, R. (2025). Personality profiles of 263 occupations. *The Journal of Applied Psychology*, *110*(4), 481–511. https://doi.org/10.1037/apl0001249

Armstrong, P. I., Allison, W., & Rounds, J. (2008). Development and initial validation of brief public domain RIASEC marker scales. *Journal of Vocational Behavior*, *73*(2), 287–299. https://doi.org/10.1016/j.jvb.2008.06.003

Barocas, S., Hardt, M., & Narayanan, A. (2023). *Fairness and machine learning: Limitations and opportunities*. MIT Press.

Batista, J. S., & Gondim, S. M. G. (2022). Personality and person-work environment fit: A study based on the RIASEC model. *International Journal of Environmental Research and Public Health*, *20*(1), Article 719. https://doi.org/10.3390/ijerph20010719

Bergmann, C. (2003). Berufliche Interessentests: Wider die Anwendung geschlechtsspezifischer Normen [Vocational interest inventories: Cons in the use of gender-specific norms]. *Zeitschrift für Personalpsychologie*, *2*(2), 66–77. https://doi.org/10.1026//1617-6391.2.2.66

Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. *PMLR Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, *81*, 149–159. https://proceedings.mlr.press/v81/binns18a.html

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Danner, D., Rammstedt, B., Bluemke, M., Lechner, C., Berres, S., Knopf, T., Soto, C. J., & John, O. P. (2019). Das Big Five Inventar 2 [The Big Five Inventory 2]. *Diagnostica*, *65*(3), 121–132. https://doi.org/10.1026/0012-1924/a000218

Danner, D., Rammstedt, B., Bluemke, M., Treiber, L., Berres, S., Soto, C. J., & John, O. P. (2016). *Die deutsche Version des Big Five Inventory 2 (BFI-2)* [*The German version of the Big Five Inventory 2*]. ZIS. https://doi.org/10.6102/zis247

Etzel, J. M., Wille, B., De Fruyt, F., & Nagy, G. (2025). Linking personality traits to vocational interest profiles via the circumplex: Research synthesis and new applications. *Journal of Vocational Behavior*, *163*. Article 104191. https://doi.org/10.1016/j.jvb.2025.104191

Gottfredson, L. S. (1982). The sex fairness of unnormed interest inventories. *Vocational Guidance Quarterly*, *31*(2), 128–132. https://doi.org/10.1002/j.2164-585X.1982.tb01303.x

Hell, B., & Buri, F. (2023). Psychometrische Überprüfung des Berufswahl- und Laufbahnplanungsinstruments Explorix [Psychometric Evaluation of the Explorix career choice and planning tool]. *Diagnostica*, *69*(1), 38–49. https://doi.org/10.1026/0012-1924/a000302

Heydasch, T. (2014). *The Hagen matrices test (HMT)* [Part of doctoral dissertation]. University of Hagen, Germany. https://doi.org/10.13140/RG.2.2.31433.75361

Hilton, M. L., & Tippins, N. T. (2010). *A database for a changing economy: Review of the Occupational Information Network (O*NET)*. National Academies Press.

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

José-García, A., Sneyd, A., Melro, A., Ollagnier, A., Tarling, G., Zhang, H., Stevenson, M., Everson, R., & Arthur, R. (2022). C3-IoC: A career guidance system for assessing student skills using machine learning and network visualisation. *International Journal of Artificial Intelligence in Education*, *33*(4), 1–28. https://doi.org/10.1007/s40593-022-00317-y

Larson, L. M., Rottinghaus, P. J., & Borgen, F. H. (2002). Meta-analyses of Big Six interests and Big Five personality factors. *Journal of Vocational Behavior*, *61*(2), 217–239. https://doi.org/10.1006/jvbe.2001.1854

Liao, H.-Y., Armstrong, P. I., & Rounds, J. (2008). Development and initial validation of public domain basic interest markers. *Journal of Vocational Behavior*, *73*(1), 159–183. https://doi.org/10.1016/j.jvb.2007.12.002

Liaw, A., & Wiener, M. (2002). Classification and regression by random Forest. *R news*, *2*(3), 18–22.

Liu, Z., Hoff, K. A., Chu, C., Oswald, F. L., & Rounds, J. (2025). Toward whole-person fit assessment: Integrating interests, values, skills, knowledge, and personality using the Occupational Information Network (O*NET). *The Journal of Applied Psychology*, *110*(5), 623–647. https://doi.org/10.1037/apl0001232

Manganello, F., Rasca, E., Villa, A., Maddalena, A., & Boccuzzi, G. (2025). Machine learning for post-diploma educational and career guidance: A scoping review in AI-driven decision support systems. *Frontiers in Education*, *10*, Article 1578979. https://doi.org/10.3389/feduc.2025.1578979

McDonald, R. P. (2013). *Test theory: A unified treatment*. Psychology Press. https://doi.org/10.4324/9781410601087

Menis–Mastromichalakis, O., Filandrianos, G., Symeonaki, M., Stamatopoulou, G., Parsanoglou, D., & Stamou, G. (2026). Gender bias in machine learning: Insights from official labour statistics and textual analysis. *Quality & Quantity*, *60*(1), 619–653. https://doi.org/10.1007/s11135-025-02261-0

Mõttus, E., & Mõttus, R. (2025). *Your personality among different jobs* [Web application]. https://apps.psych.ut.ee/JobProfiler

Neubauer, A. C., & Hofer, G. (2021). Self-estimates of abilities are a better reflection of individuals' personality traits than of their abilities and are also strong predictors of professional interests. *Personality and Individual Differences*, *169*, Article 109850. https://doi.org/10.1016/j.paid.2020.109850

R Core Team. (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rosenfelder, A., Levitin, M. D., & Gilead, M. (2025). Towards social superintelligence? AI infers diverse psychological traits from text without specific training, outperforming human judges. *Computers in Human Behavior: Artificial Humans*, *6*, Article 100228. https://doi.org/10.1016/j.chbah.2025.100228

Savickas, M. L. (2013). Career construction theory and practice. In S. D. Brown & R. W. Lent (Eds.), *Career development and counseling: Putting theory and research to work* (2nd ed., pp. 147–183). Wiley.

Schweizerisches, Dienstleistungszentrum Berufsbildung SDBB. (2025). *Einführung in die Swissdoc-Systematik* [Introduction to the Swissdoc classification system]. https://swissdoc.sdbb.ch/downloads/2025_Einfuehrung_in_die_Swissdoc.pdf

Steppan, M. (2025). *Technical report: Architecture and functionality of meinBeruf. ch*. https://meinberuf.ch/Technical%20Report.pdf

Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and things, women and people: A meta-analysis of sex differences in interests. *Psychological Bulletin*, *135*(6), 859–884. https://doi.org/7364

Wetzel, E., & Hell, B. (2013). Gender-related differential item functioning in vocational interest measurement: An analysis of the AIST-R. *Journal of Individual Differences*, *34*(3), 170–183. https://doi.org/10.1027/1614-0001/a000112

History
Received January 29, 2026
Revision received May 7, 2026
Accepted June 8, 2026
Published online July 21, 2026

Section: Methodological Topics in Assessment

Acknowledgments
An earlier version of this paper was presented at the 16th European Conference on Psychological Assessment (Barcelona, Spain, July 2025). The authors thank all participants for their valuable contributions to this study. The authors also thank the students at Macromedia University of Applied Sciences in Freiburg im Breisgau for their support with participant recruitment and data collection, as well as for providing user experience (UX) and usability feedback during the study process.

Conflict of Interest
The first author, Martin Steppan, is the owner and maintainer of meinBeruf.ch that publishes the assessment system discussed in this article. MS receives royalties related to the use of the JobMatch algorithm implemented in meinBeruf.ch.

Publication Ethics
No formal institutional review board/ethics board approval was obtained. The study used voluntary online platform/evaluation data, informed consent was obtained from all participants, and the procedures followed the ethical standards of the Swiss Psychological Society.

Authorship
Martin Steppan, conceptualization, formal analysis, methodology, data collection, software, writing – original draft, writing – review & editing; George Jogho, data collection, writing – review & editing; ZZ, formal analysis, writing – review & editing; Benedikt Hell, formal analysis, writing – review & editing.
All authors approved the final version of the article.

Funding
The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID
Martin Steppan
https://orcid.org/0000-0002-6196-9689
Benedikt Hell
https://orcid.org/0000-0002-8560-8378

Martin Steppan
Computational Psychology Lab
4056 Basel
Switzerland
mhsteppan@gmail.com