Construct Validity of Indonesian Vocational Interest Test (Tes Minat Vokasional Indonesia) to Measure Interest of Vocational School Student

Rifqa Putri Ayuni[1], Moh. Abdul Hakim[1*]
[1]Faculty of Psychology, Universitas Sebelas Maret, Surakarta, Indonesia

Abstract. This study aims to comprehensively evaluate the construct validity of Tes Minat Vokasional Indonesia (TMVI) in gauging the interests of vocational high school students. To achieve the stated objective, confirmatory factor analysis (CFA) was adopted as the study approach. Accordingly, the respondents of this investigation comprised 474 students from SMKN 2 Surakarta. The obtained results showed that model fit indicators such as CFI, TLI, SRMR, and RMSEA, all fell within the designated ranges (CFI and TLI ≥ .90, SRMR ≤ .08, and RMSEA ≤ .06). This suggest that TMVI had a rather good construct to measure the interests of vocational high school students. However, the factor loading analysis conducted reflected that some items showed factor loading values below the designated threshold (≥0.15). This observation invariably imply that the contribution of the items having low factor loading values to the measured construct was minimal. A plausible reason for this is because the vocabulary used by vocational high school pupils was not age-appropriate. Therefore, if TMVI is to be used to gauge the interests of vocational high school students, its wordings must be modified.

Keywords: Vocational high school students, vocational interest test, validity construct

Psympathic :
Jurnal Ilmiah Psikologi
Vol.12:2, December 2025,
Page 125-136

eISSN: 2502-2903
pISSN: 2356-3591

Article Info
Received:
May 02, 2025
Accepted:
December 08, 2025
Published:
December 31, 2025

DOI:
https://doi.org/10.15575/psy.v12i2.44353

*Corresponding author: Faculty of Psychology, Universitas Sebelas Maret, Surakarta, Indonesia
E-mail: m.a.hakim@staff.uns.ac.id

Introduction

Sekolah Menengah Kejuruan (SMK), or Vocational Schools, comprises secondary educational institutions that teach students specific skills such as mechanics, automotive technology, and electrical engineering, all of which are essential in the world of work. Before entering the world of work, vocational high school students need thorough preparation. An important step in this preparation is identifying the student's personal interests, as this would help the demographic choose a career path that corresponds with their skills and aspirations. As stated in a previous exploration, interest is a psychological state and motivational tendency, or the result of the interaction between individual characteristics and their environment (Renninger & Hidi, 2015). Within the context of vocational high school students, a proper understanding of one's interests is very important. This is because interest is the most prominent predictor of career success among other antecedents, such as socio-demographic background and individual capital (Su, 2020).

With a clear understanding of personal interests, a vocational high school student can identify the most suitable field of work, leading to better preparation for entering the world of work. This is very important because the suitability between interests and work significantly affects job satisfaction (Hoff et al., 2022). To identify personal interests, vocational high school students can use an interest test. This form of test has been reported to help students determine a suitable college major or field of interest. (Mahardika et al., 2024).

One of the interest tests commonly used to help vocational high school students plan respective careers is the RIASEC test developed by Holland. According to Holland's theory, the achievement of individual career satisfaction is based on the suitability between career personality and work environment (Holland, 1973). Within this context, Holland divides individual personality into six dimensions consisting of Realistic (R), Investigative (I), Artistic (A), Social (A), Enterprising (E), and Conventional (C), which are often called RIASEC.

The RIASEC model describes the six personality types in a hexagonal structure, with each dimension at one end of the hexagon. The distance between each dimension reflects the degree of their similarity (Holland, 1973). Meanwhile, Prediger (1982) suggested that interests and environments could be characterized by two dimensions, namely the

person/thing and data/idea dimensions. The Person/Thing dimension emphasizes the distinction between Social and Realistic types, which lie on opposite sides of the hexagon. On the other hand, the Data/Idea dimension distinguishes between Enterprising and Conventional types on the data side, and between Investigative and Artistic types on the idea side.

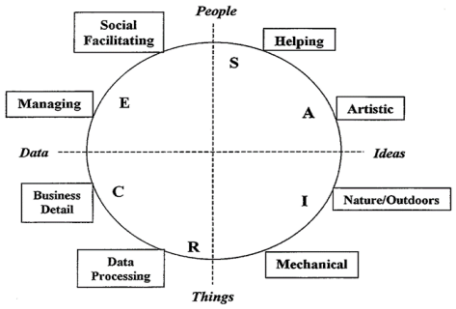

Figure 1. Basic Interest PGI (Tracy, 2002)

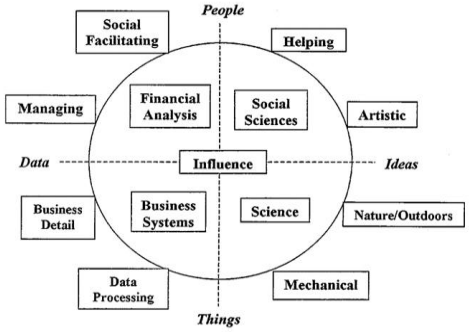

Figure 2. Higher Prestige PGI (Tracey, 2002)

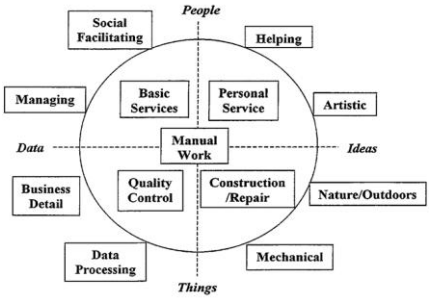

Figure 3. Lower Prestige PGI (Tracey, 2002)

The use of the Holland Test can help teachers in the career counseling process for vocational high school students (Arrasuli, 2022). For instance, in an investigation conducted by Muqaffi et al. (2022) used the Holland interest test was used to measure the interests of students, after which the obtained results were used as a reference in choosing suitable career paths for the demographic. Another study conducted at Vocational Education and Training (a type of vocational school in Germany) stated that Holland-based interest test could significantly help students find suitable fields of work (Volodina et al., 2015). Over time, the theory proposed by Holland has been used as a basis across different other investigations to develop more accurate interest tests, among which is the Personal Globe Inventory (PGI) interest test developed by (Tracey, 2002).

The Personal Globe Inventory (PGI) is an interest test developed by Tracey (2002) based on the Spherical Model of Interest theory proposed by Tracey & Rounds (1996). This model expands Holland's six personality types into eight basic interest types and introduces an additional dimension, namely prestige, which represents the level of difficulty, expertise, training, knowledge, and education required for a given activity. The investigators incorporated two prestige scales, -high and low-, in the model, after which Tracey (2002) organized 18 aspects of interest according to the dimensions of objects/people, ideas/data, and prestige.

In Indonesia, Yudiana et al. (2011), have developed the Padjadjaran Interest Inventory (PII) based on the PGI theory by Tracey (2002). Essentially, PII measures two scales, namely preferences and competencies. The interest test was reported to have good reliability as well as a valid internal structure. It can also be applied to large groups quite easily and used for career exploration (Yudiana et al., 2011). However, PII has several weaknesses, namely the geographical limitations of the sample used for the validity test, the low factor loading on several items, and the validity related to other test tools (Yudiana et al., 2011).

PII uses the Likert scale to measure interest. This scale adopts a rating system to assess the level of participant agreement with each field, and the final result is determined by summing the scores of all items (Likert, 1932). However, it is important to comprehend that the Likert scale has a limitation: participants often tend to select neutral or socially desirable responses . Based on previous observations, if a Likert scale uses answer choices that have a neutral option in the middle, participants tend to avoid extreme answer choices (strongly agree or strongly disagree) and choose options perceived as "safe" answers (Ahn & Kang, 2018; Pimentel, 2019). This is called the middle bias phenomenon and will invariably make the processing of data very difficult (Kusmaryono et al., 2022; Malone et al., 2014). The use of the Likert scale is also susceptible to social desirability, where participants will also tend to choose answers that are considered correct by society (Taherdoost, 2019). Participants will choose "strongly agree" for positive questions and "strongly disagree" for negative questions (Kusmaryono et al., 2022).

Proceeding from the limitations of the PII, which still uses the Likert scale, Hakim et al. (2024) developed the Indonesian Vocational Interest Test (TMVI), using a forced-choice measurement format. This interest test tool was also developed based on the

PGI theory introduced by Tracey (2002). Essentially, TMVI measures individual interests and self-efficacy using a forced choice (FC) scale where individuals are asked to choose one of three answer choices that are most similar to themselves. The three answer choices are arranged based on common, medium, and rare levels.

To prove the effectiveness of the FC scale compared to the Likert scale, Lee et al. (2019) conducted a study to see the differences in participant responses on the Likert and FC scales. Lee et al. (2019) asked participants to fill out the Big Five scale in two conditions: the first being the honest condition, where participants were told that the answers would only be used for study purposes, while the second condition was the faking condition, where participants were asked to imagine that they were applying for their dream job. The results of the study showed that participant responses were more stable between the honest and faking conditions using the FC scale (Lee et al., 2019). Based on this observation, it was inferred that the use of FC could ensure the quality of the decision-making process that is compiled based on the participant's profile (Lee et al., 2019).

Other studies have shown that the use of FC makes it almost impossible for respondents to support all answers in the desired manner (Brown & Maydeu-Olivares, 2018). For instance, a previous study showed that the use of the FC scale could increase criterion validity by 50% compared to the use of the Likert scale (Bartram, 2007). This scale has also been observed to be more resistant to faking and specifically demands the cognitive abilities of participants when providing answers (Zhang et al., 2024). Although the forced-choice scale type has been shown to be more effective than the Likert scale, a validity test is needed to prove whether the TMVI developed by Hakim et al. (2024) is able to measure individual interests.

In the process of developing a test instrument, one of the stages that must be carried out is the validity test. Validity is typically carried out to measure the extent which a test instrument is able to measure what should be measured (Azwar, 1987). Generally, there are three types of validity, namely content, construct, and criterion validity. Content validity simply assesses how relevant the items in an instrument measure a construct (Zamanzadeh et al., 2015). Criterion validity measures the correlation between a test instrument with other test instruments (Cohen & Swerdlik, 2005). Meanwhile, construct validity focuses on evaluating how precisely the test instrument measures what is to be measured according to the established conceptual definition (Matondang, 2009).

Construct validity can be carried out using the Confirmatory Factor Analysis (CFA), Multitrait-Multimethod, or Rasch Model analysis tests. CFA is one of the methods often used in construct validity (Umar & Nisa, 2020). The method essentially examines the relationship between observed variables or indicators (such as test items, test scores, behavioral observation ratings) and latent variables or factors (Brown, 2015). As defined in a previous exploration, latent factors are variables that cannot be directly observed but can influence multiple observed variables and are responsible for the relationships among them (Brown, 2015). The observed variables are correlated with each other because they share the same underlying factor, and if this factor is removed, the correlation between the observed variables becomes zero (Brown, 2015).

CFA is used to test the extent to which all items in a measuring instrument provide information about what is to be measured (Umar & Nisa, 2020). Within the context of this study, CFA was used to verify the number of dimensions underlying the measuring instrument (factors) and the pattern of relationships between items and factors (factor loadings) (Brown, 2015). The method uses Structural Equation Modeling (SEM), which can show the model fit index by examining the residual value, such as RMSEA. A small residual value shows a high model fit index (Raykov & Marcoulides, 2012). However, the sample used for CFA must be able to represent the population. If the sample cannot represent the population, this can lead to bias and inaccurate results (Goudarzian, 2023). For example, if the sample only includes individuals from a certain age group or socioeconomic status, then the factor structure may not be generalizable to a larger population. During the course of this study, the TMVI tool developed by Hakim et al. (2024) was tested for reliability and construct validity using CFA. The obtained results showed that TMVI is reliable and valid with some minor revisions. Regardless, the test was conducted using a sample of adults with work experience. To determine whether TMVI can accurately measure the interests of vocational high school students, a validity test should be conducted with an appropriate student sample.

In addition to CFA, another model called the Multiple Indicators Multiple Causes (MIMIC) model was used to determine whether the instrument is influenced by individual characteristics. As stated in a prior study, the MIMIC model essentially attempts to accommodate population heterogeneity by incorporating a set of predictors or covariates into the model (Muthén, 1989). Generally, instruments that can accommodate such diversity in population characteristics are considered to have strong external validity (Widhiarso, 2012). This approach complements the primary purpose of the present study,

Table 1
Blueprint of Tes Minat Vokasional Indonesia (TMVI)

Interest Area	Interest Aspect	Item Number	Total Item
Basic Interest	Social Facilitating	SF1, SF2, SF3, SF4, SF5, SF6	6
	Managing	MG1, MG2, MG3, MG4, MG5, MG6	6
	Business Detail	BD1, BD2, BD3, BD4, BD5, BD6	6
	Data Processing	DP1, DP2, DP3, DP4, DP5, DP6	6
	Mechanical	MC1, MC2, MC3, MC4, MC5, MC6	6
	Nature/Outdoors	NTO1, NTO2, NTO3, NTO4, NTO5, NTO6	6
	Artistic	ART1, ART2, ART3, ART4, ART5, ART6	6
	Helping	HLP1, HLP2, HLP3, HLP4, HLP5, HLP6	6
Higher Prestige	Social Science	SS1, SS2, SS3, SS4, SS5, SS6	6
	Influence	INF1, INF2, INF3, INF4, INF5, INF6	6
	Business System	BS1, BS2, BS3, BS4, BS5, BS6	6
	Financial Analysis	FIN1, FIN2, FIN3, FIN4, FIN5, FIN6	6
	Science	SC1, SC2, SC3, SC4, SC5, SC6	6
Lower Prestige	Quality Control	QC1, QC2, QC3, QC4, QC5, QC6	6
	Basic Service	BSC1, BSC2, BSC3, BSC4, BSC5, BSC6	6
	Personal Service	PS1, PS2, PS3, PS4, PS5, PS6	6
	Manual Work	MW1, MW2, MW3, MW4, MW5, MW6	6
	Construction/Repair	CR1, CR2, CR3, CR4, CR5, CR6	6
Total Item			108

Table 2
Example of Tes Minat Vokasional Indonesia (TMVI)

Interest Aspect	Level	Item Question
Social Facilitating	Common	Asking about other people's needs.
	Medium	Prioritize the interests of others even in difficult situations.
	Rare	Willing to volunteer for any situation.

which is to evaluate the construct validity of the Tes Minat Vokasional Indonesia (TMVI) in measuring the interests of vocational high school students.

Methods

Participants

The present study's population comprised 2423 students from SMKN 2 Surakarta (Data Pokok Pendidikan Kementerian Pendidikan Dasar dan Menengah, 2024). However, 474 of the students, all of whom were between grades 10 to 12, served as the sample. The participants were from 6 different majors, including electrical engineering, mechanical engineering, geomatics engineering, electronic engineering, welding & metal fabrication engineering, as well as computer network & telecommunication engineering. The sample represents approximately 19.6% of the total population, and the adequacy of its size is supported by Krejcie and Morgan's (1970) sample size table, which recommends a minimum of 331 respondents for a population of around 2,400. Therefore, the sample of 474 students exceeds the required minimum and is considered sufficient. The sampling method adopted was convenience sampling, where data were obtained from participants who were easily accessible (Rahi, 2017). After consulting with the school, participants were selected based on their availability during the data collection period. Students who were still attending classes, rather than those already engaged in internships, were selected to take part in this study. After the selection process, the TMVI interest test tool was distributed via a Google form filled out by students under direct supervision to mitigate errors.

Table 3
Model Fit Analysis

Interest	CFI	TLI	SRMR	RMSEA
SF	.965	.934	.026	.023
MG	1.000	2.400	.013	.000
BD	.942	.903	.025	.028
DP	1.000	1.010	.020	.000
MC	1.000	1.020	.021	.000
NTO	.987	.976	.024	.024
ART	.989	.981	.024	.020
HLP	1.000	1.010	.021	.000
SS	.977	.950	.024	.031
INF	.949	.916	.037	.032
BS	.983	.971	.025	.022
FIN	.969	.942	.025	.026
SC	1.000	1.000	.019	.000
QC	1.000	1.000	.022	.000
BSC	1.000	1.020	.023	.000
PS	.975	.947	.023	.029
MW	.982	.967	.025	.021
CR	1.000	1.010	.021	.000

Instrument

The instrument used in this study was the Indonesian Vocational Interest Test (TMVI) developed by Hakim et al. (2024). This test assesses both individual interests and self-efficacy. It uses a forced-choice format, where participants select one of three options categorized as common, medium, or rare. Accordingly, the instrument evaluates individual interests across 18 aspects, with each aspect comprising six items. (see Table 1). Table 2 presents some of these items.

Data Analysis

The analysis method adopted in this study is CFA. CFA was carried out with the help of Jamovi 2.6.13 and JASP. The factor loading indicator was used to assess the validity of the TMVI measuring instrument, with a loading of ≥ .15 considered acceptable for a sample size exceeding 300 participants (Kline, 2014). Other indicators used in the CFA analysis include the Tucker-Lewis Index (TLI), Comparative Fit Index (CFI), Standardized Root Mean Square (SRMR), and Root Mean Squared Error of Approximation (RMSEA). The fit criteria for each indicator include TLI ≥ .90, CFI ≥ .90, SRMR ≤ .08, and RMSEA ≤ .06 (Wang & Wang, 2020). An additional analysis using the MIMIC model was conducted to determine whether participants' interests were influenced by external factors such as age, gender, grade level, or major, with correlations above .05 showing no significant effect.

Results and Discussion

Results

The results of the model fit analysis (see Table 3) show that all aspects had a good match with the data. This can be seen from the obtained CFI, TLI, SRMR, and RMSEA values, which were in the acceptable range, with CFI and TLI ≥ .90, SRMR ≤ .08, and RMSEA ≤ .06 (Wang & Wang, 2020).

Table 4
Factor Loading Analysis Basic Interest

Interest	Item	Est	S.E.	p
Social Facilitating	SF1	.223	.054	.007
	SF2	.198	.067	.014
	SF3	.357	.065	≤ .001
	SF4	.434	.068	≤ .001
	SF5	.211	.053	.014
	SF6	.261	.045	.003
Managing	MG1	.231	.800	.015
	MG2	.262	.598	.007
	MG3	.173	.577	.050
	MG4	.420	.111	.002
	MG5	.017	.060	.843
	MG6	.204	.074	.027
Business Detail	BD1	.073	.055	.310
	BD2	.298	.057	≤ .001
	BD3	.217	.055	.008
	BD4	.323	.072	≤ .001
	BD5	.170	.045	.016
	BD6	.573	.088	≤ .001
Data Processing	DP1	.448	.056	≤ .001
	DP2	.330	.057	≤ .001
	DP3	.361	.060	≤ .001
	DP4	.365	.059	≤ .001
	DP5	.329	.058	≤ .001
	DP6	.244	.056	≤ .001
Mechanical	MC1	.391	.063	≤ .001
	MC2	.216	.041	.003
	MC3	.486	.049	≤ .001
	MC4	.260	.059	≤ .001
	MC5	.293	.608	≤ .001
	MC6	.293	.046	≤ .001
Nature/ Outdoors	NTO1	.411	.053	≤ .001
	NTO2	.176	.042	.004
	NTO3	.618	.061	≤ .001
	NTO4	.441	.045	≤ .001
	NTO5	.406	.039	≤ .001
	NTO6	.301	.054	≤ .001
Artistic	ART1	.502	.027	≤ .001
	ART2	.253	.037	≤ .001
	ART3	.472	.037	≤ .001
	ART4	.303	.036	≤ .001
	ART5	.441	.035	≤ .001
	ART6	.492	.031	≤ .001
Helping	HLP1	.098	.053	.095
	HLP2	.308	.043	≤ .001
	HLP3	.572	.051	≤ .001
	HLP4	.578	.042	≤ .001
	HLP5	.564	.049	≤ .001
	HLP6	.230	.039	≤ .001

Beyond assessing model fit, factor loading analysis was adopted to evaluate the contribution of each item within its respective aspect. Item with high factor loading values were considered to effectively

measure the intended latent construct. In this study, a factor loading of ≥ .15 was used as the threshold for acceptability (Kline, 2014). The results of the analysis show that in the basic interest dimension (see Table 4), the Managing, Business Detail, and Helping aspects each had one item below the minimum limit, namely MG5, BD1, and HLP1, respectively. Similarly, in the Higher interest dimension (see Table 5), the Business System aspect had 1 item below the minimum limit, namely BS1. Within the Lower interest dimension (see Table 6), the Construction/Repair aspect also had 1 item below the minimum limit, namely CR1.

Table 5
Factor Loading Analysis: Higher Prestige

Interest	Item	Est	S.E.	p
Social Science	SS1	.251	.050	≤ .001
	SS2	.318	.053	≤ .001
	SS3	.266	.053	≤ .001
	SS4	.452	.054	≤ .001
	SS5	.426	.053	≤ .001
	SS6	.483	.045	≤ .001
Influence	INF1	.395	.046	≤ .001
	INF2	.219	.082	.030
	INF3	.394	.057	≤ .001
	INF4	.344	.054	≤ .001
	INF5	.389	.046	≤ .001
	INF6	.411	.053	≤ .001
Business System	BS1	.114	.039	.071
	BS2	.433	.045	≤ .001
	BS3	.320	.051	≤ .001
	BS4	.507	.054	≤ .001
	BS5	.435	.044	≤ .001
	BS6	.424	.048	≤ .001
Financial Analysis	FIN1	.303	.056	≤ .001
	FIN2	.367	.054	≤ .001
	FIN3	.472	.064	≤ .001
	FIN4	.300	.066	≤ .001
	FIN5	.367	.048	≤ .001
	FIN6	.314	.053	≤ .001
Science	SC1	.426	.041	≤ .001
	SC2	.315	.031	≤ .001
	SC3	.569	.042	≤ .001
	SC4	.507	.036	≤ .001
	SC5	.509	.039	≤ .001
	SC6	.561	.040	≤ .001

Reliability testing was conducted to assess the instrument's internal consistency both before and after the CFA. Within this context, the obtained result showed that McDonald's ω for each interest area ranged from .20 to .65 (see Table 7). This implies that the TMVI possessed poor internal consistency, with little difference in reliability observed before and after the CFA.

The result from the MIMIC model analysis showed that the suspected predictors (age, gender, grade, and major) all had significance values above .05 (see Table 8). This implies that the suspected predictors did not affect the interest of the participant. As shown in Table 9, the suspected predictors did not have a good model fit, as reflected by CFI and TLI values below the acceptable threshold of ≥ .90 and an RMSEA value above the recommended ≤ .06, although the SRMR value of ≤.08 remained within the acceptable range (Wang & Wang, 2020).

Table 6
Factor Loading Analysis: Lower Prestige

Interest	Item	Est	S.E.	p
Quality Control	QC1	.443	.045	≤ .001
	QC2	.557	.046	≤ .001
	QC3	.479	.046	≤ .001
	QC4	.439	.045	≤ .001
	QC5	.345	.048	≤ .001
	QC6	.274	.046	≤ .001
Basic Service	BSC1	.368	.054	≤ .001
	BSC2	.281	.063	≤ .001
	BSC3	.316	.057	≤ .001
	BSC4	.353	.055	≤ .001
	BSC5	.346	.056	≤ .001
	BSC6	.287	.055	≤ .001
Personal Service	PS1	.466	.046	≤ .001
	PS2	.429	.045	≤ .001
	PS3	.213	.052	.002
	PS4	.304	.058	≤ .001
	PS5	.423	.052	≤ .001
	PS6	.347	.054	≤ .001
Manual Working	MW1	.354	.047	≤ .001
	MW2	.242	.047	≤ .001
	MW3	.401	.047	≤ .001
	MW4	.339	.051	≤ .001
	MW5	.454	.054	≤ .001
	MW6	.391	.060	≤ .001
Construction/ Repair	CR1	.109	.047	.068
	CR2	.447	044	≤ .001
	CR3	.488	.046	≤ .001
	CR4	.482	.041	≤ .001
	CR5	.588	.047	≤ .001
	CR6	.320	.043	≤ .001

Discussion

This study aims to test the construct validity of the Indonesian Vocational Interest Test (TMVI) in measuring the interests of vocational high school students. To achieve the stated aim, CFA analysis was conducted, and the model fit was evaluated using the CFI, TLI, SRMR, and RMSEA indicators. The results obtained from the model fit analysis (see Table 3) showed that all aspects of the model had CFI, TLI, SRMR, and RMSEA values within the acceptable range, with CFI and TLI ≥ .90, SRMR ≤ .08, and RMSEA ≤ .06 (Wang & Wang, 2020). This reflects that all interest types measured in the TMVI showed adequate construct validity.

SRMR and RMSEA were included in the absolute fit indices, which evaluated the degree of correspondence between the specified model and the

sample data (Hair et al., 2010). Based on predefined standards, low SRMR and RMSEA values reflect that there is only a slight difference between the model and the data; hence, the established model would be considered fit (Hair et al., 2010). Within the context of this study, the SRMR and RMSEA values (see Table 3) for all types of interest in TMVI fell below the predetermined cut-off limits, namely SRMR≤ .08 and RMSEA≤ .06 (Wang & Wang, 2020). These results showed that the determined model was in accordance with the data obtained.

Table 7
Reliability before and after CFA

Interest Area	McDonald's ω Before CFA	McDonald's ω After CFA
SF	.351	.374
MG	.239	.238
BD	.344	.343
DP	.404	.400
MC	.455	.452
NTO	.553	.552
ART	.535	.550
HLP	.528	.533
SS	.516	.473
INF	.465	.435
BS	.505	.498
FIN	.440	.445
SC	.653	.646
QC	.565	.567
BSC	.411	.415
PS	.473	.480
MW	.453	.454
CR	.546	.548

CFI and TLI were included in the incremental fix indices, which compared the proposed model with a null model that assumes no correlations among the observed variables (Hair et al., 2010). According to predefined standards, a model is considered fit when it shows a substantial difference from the null model, reflecting that meaningful correlations exist among the observed variables (Hair et al., 2010). CFI is an index that already has a norm and is in the range of 0 – 1, where the closer to 1, the fitter the model (Brown, 2015). Meanwhile, TLI does not have a norm and can be below 0 or above 1 (Brown, 2015). The results of this study showed that all types of interests had CFI and TLI values (see Table 3) above the set cut-off limit, which is ≥ .90 (Wang & Wang, 2020). This signified that the proposed model differed significantly from the null model, and correlations were present among the observed variables. Several other aspects were also observed to have TLI values of more than 1, namely Managing, Business Detail, Data Processing, Mechanical, Basic Service, and Construction/Repair (see Table 3). As stated in a previous investigation, a TLI value above 1 shows a very good model fit (Goretzko et al., 2024).

Table 8
MIMIC Model Analysis

Predictor	Est	S.E.	z-value	p-value
Age	.043	.050	.865	.387
Gender	.291	.174	1.672	.095
Grade	.093	.056	1.661	.097
Major	-.012	.025	-.490	.624

Table 9
MIMIC Model Fit

Predictor	CFI	TLI	SRMR	RMSEA
Age	.874	.858	.048	.119
Gender	.873	.857	.049	.120
Grade	.873	.857	.048	.120
Major	.875	.859	.048	.119

In addition to assessing model fit, CFA also provided factor loading values, which were used to determine the strength of each item's contribution to its respective aspect. Within this context, a factor loading of ≥ .15 is considered acceptable (Kline, 2014). Based on the CFA results, several TMVI items showed very low factor loading values. In the basic interest dimension (see Table 4), the Managing aspects, Business Detail, and Helping aspects each had one item below the minimum threshold, namely MG5, BD1, and HLP1. Accordingly, in the higher interest dimension (see Table 5), the Business System aspect had one item below the minimum threshold (BS1). In the lower interest dimension (see Table 6), the Construction/Repair aspect similarly had one item below the minimum threshold (CR1). Because these items had factor loading values below 0.15 (Kline, 2014), each was interpreted as having a weak contribution to the respective aspects and was therefore considered for deletion. After the items were omitted, all remaining items showed good factor loading values, reflecting that each included item was capable of measuring the intended construct effectively.

One example of an item with a very low factor loading value is BS1, which included the following response options: a common-level choice of "investing in stocks," a medium-level choice of "learning stock investment strategies," and a rare-level choice of "taking stock investment training." This item contributes minimally to the Business Detail interest. A plausible explanation is the wording used, particularly vocabulary such as investment and stocks, which may be difficult for vocational school students to understand. After data collection, participants were asked about their testing experience, and several students reported encountering unfamiliar terms that affected their comprehension of certain questions. Based on the observations made, inappropriate or overly complex language can limit an instrument's

ability to measure the intended construct, making it essential to ensure that the wording corresponds with the characteristics of the target population (Hambleton et al., 2004). Since the TMVI was originally developed to assess the interests of adult employees, the language used in the instrument must be adjusted if it is intended to measure the interests of vocational school students.

The reliability analysis using McDonald's ω (see Table 7) showed that the TMVI did not reflect strong internal consistency, with minimal differences observed between the reliability values obtained before and after the CFA. An additional analysis was carried out using the MIMIC model to examine whether the suspected predictors could influence the interests of participants (Widhiarso, 2012). The obtained results showed that the predictors, namely age, gender, grade, and major, did not correlate with the interest scores (see Table 8). This result shows how the TMVI measured only the interests of the participants without being affected by external factors, signifying that the instrument can be applied to a more heterogeneous population.

In entirety, all types of interests measured by the TMVI showed adequate construct validity for assessing the interests of vocational high school students, as reflected by model fit indices including CFI, TLI, RMSEA, and SRMR. However, from the factor loading, several items were observed to only provide a small contribution to the measured construct. This minimal contribution can be attributed to the use of language that is less appropriate for the target population of students. Based on this understanding, it becomes necessary to make adjustments to TMVI in order for the model to effectively measure the interests of vocational high school students.

The advantage of TMVI compared to the Holland Test, which is often used to measure student interests, is that TMVI uses a forced-choice scale. This is different from the Holland test, which still uses the Likert scale. The significance within this context includes the fact that the forced-choice scale is able to produce a more stable personality profile than the Likert scale, is more resistant to biased responses, and demands the cognitive abilities of participants when working on answers (Lee et al., 2019; Zhang et al., 2024). However, TMVI was observed to possess a significant weakness in this study, which includes the use of language that is less appropriate for vocational high school students. As stated in a prior investigation, the use of language that is less appropriate to the target population can significantly affect the ability of the test tool to effectively measure the intended construct (Hambleton et al., 2004). Therefore, if TMVI will be used to measure the interest of vocational students, the language on some items that have very low factor loading has to be modified to facilitate the understanding of students and enable the demographic comprehend the items better and answer more correctly. Based on observation, modifying the stated items would make TMVI more valid and reliable for the measurement of vocational students' interest. Furthermore, the model could then be used by guidance and counseling teachers to assist students in making decisions about respective majors and career paths that best suit the interests and abilities of each individual.

Despite the results above, this study is limited to examining the construct validity of the TMVI using CFA. Regardless of the fact that the analysis supported the internal structure of the instrument, it did not assess how well the TMVI relates to external variables or outcomes. To strengthen the entire validity and practical utility of the instrument, future explorations should investigate criterion validity by comparing TMVI scores with external measures, such as clinical assessments, behavioral evaluations, or other established psychological tests.

Conclusion

In conclusion, the results obtained from this study showed that the Indonesian Vocational Interest Test (TMVI) possessed adequate construct validity to measure the interests of vocational high school students in terms of model fit indicators such as CFI, TLI, SRMR, and RMSEA, all of which were within the specified limits (CFI and TLI $\geq .90$, SRMR $\leq .08$, and RMSEA $\leq .06$). However, the results of the factor loading analysis reflected that various items had factor loading values below the specified limit ($\geq .15$). These items were found to have minimal contribution to the measured construct, possibly due to the use of language that was not well-suited for vocational high school students. Based on this insight, it is necessary to adjust the language of the TMVI before using the model to measure the interests of vocational high school students.

Declaration

Acknowledgments

This researcher is grateful for the assistance from Mr. Moh. Abdul Hakim as the supervisor for this study, teachers and students at SMKN 2 Surakarta who are involved in the research, and Faculty of Psychology, Sebelas Maret University that facilitated this study.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contributions

Conceptualization: R.P.A; Methodology: M.A.H; Data collection and investigation: R.P.A; Data analysis:

R.P.A; Writing—Original Draft Preparation: R.P.A; Writing—Review & Editing: M.A.H; Supervision: M.A.H.

Conflict of Interest

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

Use of Artificial Intelligence

The authors declare that no Artificial Intelligence (AI) or AI-assisted technologies were used in the creation of this manuscript.

Ethical Clearance

Ethical review and approval were waived for this study as there are no sensitive variable involved and did not cause harm to the respondents.

Data Availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

References

Ahn, E., & Kang, H. (2018). Introduction to systematic review and meta-analysis. *Korean Journal of Anesthesiology*, 71(2), 103–112. https://doi.org/10.4097/kjae.2018.71.2.103

Arrasuli, A. N. (2022). Teori Holland Untuk Membantu Pemilihan Jurusan Pada Siswa SMK. *Education & Learning*, 2(2), 88–91. https://doi.org/10.57251/el.v2i2.390

Azwar, S. (1987). *Reliabilitas dan Validitas*.

Bartram, D. (2007). Increasing Validity with Forced-Choice Criterion Measurement Formats. *International Journal of Selection and Assessment*, 15(3), 263–272. https://doi.org/10.1111/j.1468-2389.2007.00386.x

Brown, A., & Maydeu-Olivares, A. (2018). Modelling Forced-Choice Response Formats. *In The Wiley Handbook of Psychometric Testing* (pp. 523–569). Wiley. https://doi.org/10.1002/9781118489772.ch18

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford publications.

Cohen, R. J., & Swerdlik, M. E. (2005). Psychological testing and assessment : an introduction to tests and measurement. *In Guidebook for Clinical Psychology Interns*. McGraw-Hill. https://doi.org/10.1007/978-1-4899-0222-1_8

Data Pokok Pendidikan Kementerian Pendidikan Dasar dan Menengah. (2024). SMKN 2 Surakarta. https://dapo.dikdasmen.go.id/sekolah/04067A1AFFD3F4D6C5CB

Goretzko, D., Siemund, K., & Sterner, P. (2024). Evaluating Model Fit of Measurement Models in Confirmatory Factor Analysis. *Educational and Psychological Measurement*, 84(1), 123–144. https://doi.org/10.1177/00131644231163813

Goudarzian, A. H. (2023). Challenges and recommendations of exploratory and confirmatory factor analysis: A narrative review from a nursing perspective. *Journal of Nursing Reports in Clinical Practice*, 1(3), 133–137. https://doi.org/10.32598/JNRCP.23.44

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2010). *Multivariate Data Analysis* (7th ed.). John Wiley & Sons Ltd.

Hakim, UNS, & PLN. (2024). *Technical Book Tes Minat Dewasa Indonesia* (TMVI).

Hambleton, R. K., Merenda, P. F., & Spielberger, C. D. (2004). *Adapting educational and psychological tests for cross-cultural assessment*. Psychology Press.

Hoff, K. A., Chu, C., Einarsdóttir, S., Briley, D. A., Hanna, A., & Rounds, J. (2022). Adolescent vocational interests predict early career success: Two 12-year longitudinal studies. *Applied Psychology*, 71(1), 49–75. https://doi.org/10.1111/apps.12311

Holland, J. L. (1973). Making vocational choices: A theory of vocational personalities and work environments. *Psychological Assessment Resources*.

Kline, P. (2014). *An Easy Guide to Factor Analysis*. Routledge. https://doi.org/10.4324/9781315788135

Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and Psychological Measurement*, 30(7).

Kusmaryono, I., Wijayanti, D., & Maharani, H. R. (2022). Number of Response Options, Reliability, Validity, and Potential Bias in the Use of the Likert Scale Education and Social Science Research: A Literature Review. *International Journal of Educational Methodology*, 8(4), 625–637. https://doi.org/10.12973/ijem.8.4.625

Lee, P., Joo, S. H., & Lee, S. (2019). Examining stability of personality profile solutions between Likert-type and multidimensional forced choice measure. *Personality and Individual Differences*, 142 (August 2018), 13–20. https://doi.org/10.1016/j.paid.2019.01.022

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*.

Mahardika, R., & Nurhayati, S. (2024). Pemanfaatan Hasil Tes Minat bagi Guru Bimbingan dan

Konseling di Sekolah. *Jurnal Penelitian Pendidikan Indonesia (JPPI)*, *1*(2), 292-297.

Malone, H., Nicholl, H., & Tracey, C. (2014). Awareness and minimisation of systematic bias in research. *British Journal of Nursing*, 23(5), 279–282. https://doi.org/10.12968/bjon.2014.23.5.279

Matondang, Z. (2009). Validitas dan reliabilitas suatu instrumen penelitian. *Jurnal Tabularasa*, 6(1), 87–97.

Muqaffi, R., Rahmah, N., & Asis, F. (2022). Tes Minat Bakat Holland Personality Test terhadap Siswa di SMKN 3 Soppeng. *Jurnal Lepa-Lepa Open*, 2(2), 286–295.

Muthén, B. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, 54(14). https://doi.org/https://doi.org/10.1007/BF02296397

Pimentel, J. L. (2019). Some Biases in Likert Scaling Usage and its Correction. *International Journal of Sciences Basic and Applied Research*, 45(1), 183–191. http://gssrr.org/index.php?journal=JournalOfBasicAndApplied

Prediger, D. J. (1982). Dimensions underlying Holland's hexagon: Missing link between interests and occupations? *Journal of Vocational Behavior*, 21(3), 259–287. https://doi.org/10.1016/0001-8791(82)90036-7

Rahi, S. (2017). Research Design and Methods : A Systematic Review of Research Paradigms, Sampling Issues and Instruments Development. *International Journal of Economics*. 6(2). https://doi.org/10.4172/2162-6359.1000403

Raykov, T., & Marcoulides, G. A. (2012). A First Course in Structural Equation Modeling. *Routledge*. https://doi.org/10.4324/9780203930687

Renninger, K. A., & Hidi, S. E. (2015). The Power of Interest for Motivation and Engagement. *Routledge*. https://doi.org/10.4324/9781315771045

Su, R. (2020). The three faces of interests: An integrative review of interest research in vocational, organizational, and educational psychology. *Journal of Vocational Behavior*, 116, 103240. https://doi.org/10.1016/j.jvb.2018.10.016

Taherdoost, H. (2019). What Is the Best Response Scale for Survey and Questionnaire Design. *International Journal of Academic Research in Management (IJARM),* 8(1), 1–10. https://ssrn.com/abstract=3588604

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

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

Umar, J., & Nisa, Y. F. (2020). Uji Validitas Konstruk dengan CFA dan Pelaporannya. *Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia (JP3I)*, 9(2), 1–11. https://doi.org/10.15408/jp3i.v9i2.16964

Volodina, A., Nagy, G., & Köller, O. (2015). Success in the first phase of the vocational career: The role of cognitive and scholastic abilities, personality factors, and vocational interests. *Journal of Vocational Behavior*, 91, 11–22. https://doi.org/10.1016/j.jvb.2015.08.009

Wang, J., & Wang, X. (2020). *Structural Equation Modeling: Applications Using Mplus* (2nd ed.). John Wiley & Sons Ltd.

Widhiarso, W. (2012). Penerapan Model MIMIC untuk Menguji Konsistensi Hasil Pengukuran melalui Skala. 6(1), 109–118.

Yudiana, W., Purwono, U., & Wiyono, S. (2011). Technical Manual Padjadjaran Interest Iventory. January 2011, 67. https://www.researchgate.net/publication/326292604_Technical_Manual_Padjadjaran_Interest_Inventory

Zamanzadeh, V., Ghahramanian, A., Rassouli, M., Abbaszadeh, A., Alavi-Majd, H., & Nikanfar, A.-R. (2015). Design and Implementation Content Validity Study: Development of an instrument for measuring Patient-Centered Communication. *Journal of Caring Sciences*, 4(2), 165–178. https://doi.org/10.15171/jcs.2015.017

Zhang, B., Luo, J., & Li, J. (2024). Moving beyond Likert and traditional forced-choice scales: A comprehensive investigation of the graded forced-choice format. *Multivariate Behavioral Research*, 59(3), 434-460. https://doi.org/https://doi.org/10.1080/00273171.2023.2235682

Appendix

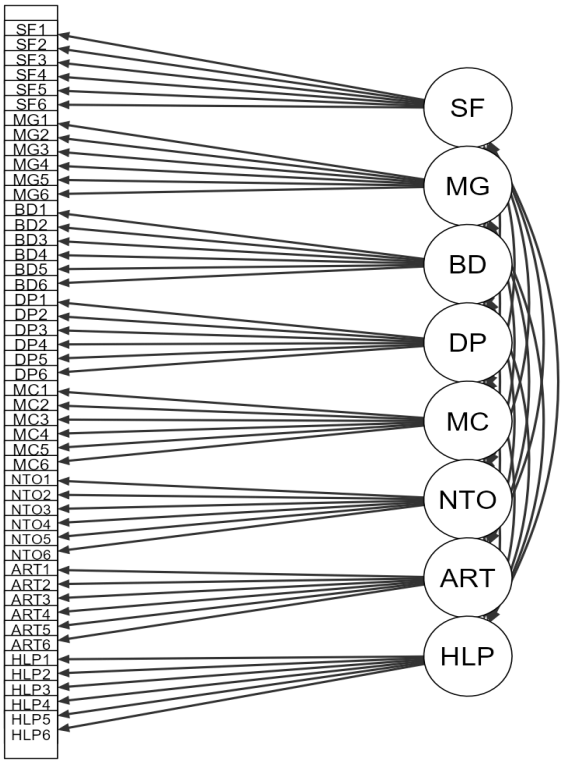

Figure 4. Basic Interest Model Diagram

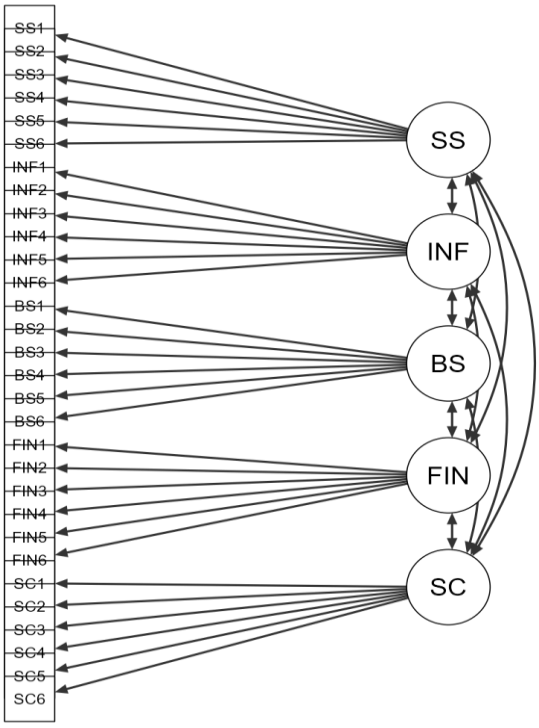

Figure 5. Higher Interest Model Diagram

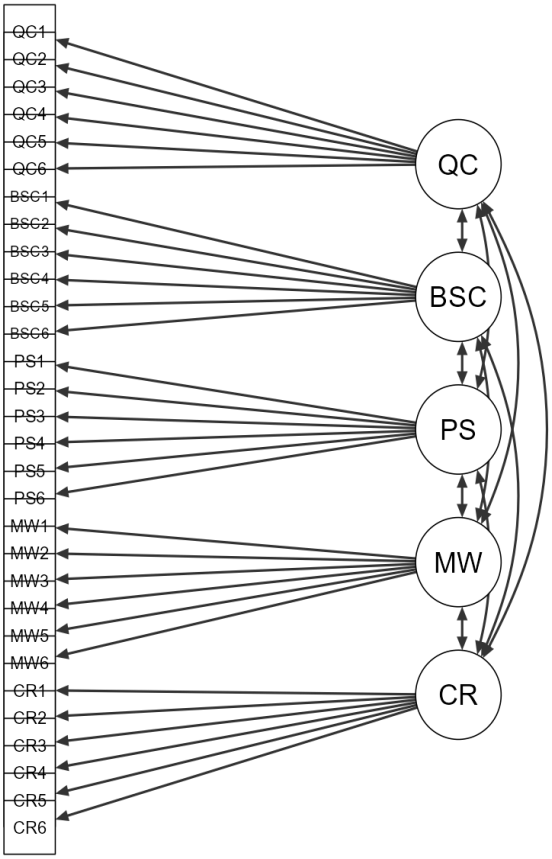

Figure 6. Lower Interest Model Diagram

This page has been intentionally left blank