JIRSEA
JOURNAL OF INSTITUTIONAL RESEARCH SOUTH EAST ASIA

JIRSEA Issue: Vol. 22 No. 3
Sept/Oct 2024
ISSN 1675-6061
http://www.seaairweb.info/journal/index.aspx

Developing Vocational Internship Program Evaluation Instrument

Yolandaru Septiana, Tony Wijaya, Edi Istiyono, and Sukirno

CITE ARTICLE: Septiana, Y., Wijaya, T., Edi Istiyono, E., and Sukirno. (2024). *Journal of Institutional Research South East Asia*, 22(3), 177-191

Submission Timeline

First submission: 8 December 2023

Revised submission: 18 July 2024

Final Submission: 14 October 2024

Acceptance: 29 October 2024

SEAAIR
SOUTH EAST ASIAN ASSOCIATION FOR INSTITUTIONAL RESEARCH

Publisher: SEAAIR Secretariat

C/O Suan Dusit Rajaphat University
295 Nakhon Ratchasima Rd, Dusit, Dusit District, Bangkok 10300, THAILAND
email:seaair.info@gmail.com
http://www.seaairweb.info/

Developing Vocational Internship Program Evaluation Instrument

Yolandaru Septiana[1], Tony Wijaya[1], Edi Istiyono[1], Sukirno[1]

[1]*Universitas Negeri Yogyakarta, Yogyakarta, Indonesia*

ABSTRACT

The internship program is momentous in providing practical experience for vocational school accounting students in Indonesia. In practice, however, different evaluation models assess the program's effectiveness. As a sequence, therefore, this study was intended to develop standardized instruments that can be adopted in the internship program evaluation at the vocational school. Initially, 31 items for the internship program evaluation were drafted. Next, by adopting the Aiken validity model, three experts were requested to assess the instrument's feasibility. Besides, an empirical validation was conducted by inviting 180 vocational school students. Finally, by using confirmatory factor analysis (CFA), root mean squared error of approximation (RMSEA), Tucker Lewis Index (TLI), and comparative fit index (CFI), this research has generated a reliable instrument consisting of 23 valid items which can be used as a standard in evaluating the internship program effectiveness for vocational school in Indonesia. The results estimate that 23 valid instruments can be used to evaluate the program of accounting field practice in vocational schools. This tool is effective in evaluating accounting fieldwork practices.

Keywords: Internship, Vocational School, Evaluation, Confirmatory Factor Analysis

Introduction

For students in vocational schools, an internship is a learning process that is carried out through work experience in the working world over a set period per the curriculum and the needs of the working world (*Regulation of the Minister of Education and Culture of the Republic of Indonesia No. 50 of 2020*, n.d.). It is important to provide practical experience for students with the competence of accounting expertise of the vocational school. The internship activity is one of the implementations of the link-and-match policy in vocational schools. It aims to improve the relevance of vocational schools to the needs of the working, business, and industry (Disas, 2018). Thus, the link and match policy is hoped to reduce the unemployment rate of vocational schools (Disas, 2018). The link and match policy is like a dual system in Germany. Dual-system training occurs in educational institutions and workplaces (Remington, 2018). The dual system in Germany is mandatory, unlike other countries, such as the UK, which is voluntary (Deissinger, 2010).

The dual system is a structural and didactic pattern in professional learning. It can be implemented in 3 ways: 1) by attending a full-time internship in a school (professional), college, or higher educational institution that does not have a training or employment contract; 2) by acquiring special skills or competencies in a company which has contract work (training participant or employee) and therefore in a particular work environment (studying while working); or 3) participating in industrial work practices or apprenticeship programs, which use a school-based (part-time) and company-based learning mode (Deissinger, 2010). The dual system aims to offer engaging training for young adults with skills who have completed three and half years of training (Haasler, 2020). However, it becomes challenging for low-achieving young people (including immigrants and refugees) to enroll in vocational and educational training (VET) programs. Germany and Switzerland are countries that have been successful in implementing VET (Deissinger & Gonon, 2021). Germany has a robust dual internship system (Haasler, 2020). The internship is being rebuilt in Germany and Switzerland, leading to the development of a trustworthy institutional internship framework and contemporary vocational training system that emphasizes employability skills (Deissinger & Gonon, 2021). Part-time vocational schools are the most prevalent in the two-country distribution system, and they are now recognized as an institution that not only links the spheres of job and education but is also backed by a solid cross-class consensus in both countries (Deissinger & Gonon, 2021). Unlike Germany, the latest internship model in the UK is STEM, where interns work while learning to earn a degree.

In Indonesia, the link and match policy is the policy of linking (linking) the competence of graduates of education according to the demand and needs of the business and industrial world in terms of quantity, quality, range, qualification, and time of internship (Maulina & Yoenanto, 2022). In the era of Industry 4.0, vocational education is a real link and match and has a huge chance (Sudira, 2019). The existence of link-and-match programs is very helpful in establishing vocational school cooperation with the business and industrial world in the form of competency-based training (CBT), the MoU program, curriculum coordination, industrial work practice, and competence test expertise (Maulina & Yoenanto, 2022).

The internship is one of the policies of link and match, a learning model that combines institutional-based education and work-based training (Astuti et al., 2023) and assists in professional growth (Anjum, 2020). The participants' accounting students are better

equipped because of the internship experience. Work readiness can also be developed and influenced by experience, the educational setting, the pedagogical expertise of the instructors, and their performance in activities related to internships (Saputra & Sukirno, 2020). This activity is followed by vocational school students, one of whom is accounting. The internship is organized in the working world, including the business world, the industry world, state-owned enterprises or regionally owned enterprises, government agencies, or other agencies. At the implementation stage of the industrial work practice, students are placed based on competence, work practice, and mentoring with the internship tutor (*Regulation of the Minister of Education and Culture of the Republic of Indonesia No. 50 of 2020*, n.d.). In addition to the school program, the internship is also organized from the Corporate Social Responsibility (CSR) program of oil and gas manufacturing companies in Indonesia, where the results are competent staff, the characteristics of participants reflect the goals of the program, a synchronous and flexible curriculum, adequate facilities, and fulfilled procedures (Ramadhani & Rahayu, 2020). There are still students whose internship does not match their competence. The level of participation of the business and industrial world in internship activities, the level of involvement in curriculum validation activities, and the participation in recruitment activities still appear insufficient (Ismail, 2022). Students at vocational schools also learn theoretical but less relevant to the needs of the labor market (Suharno et al., 2020).

The promising program still needs evaluation to improve. The study that developed evaluation instrument is from audio video engineering expertise but not for accounting, which showed the product of instruments consisting of context evaluation (based on weaknesses, strengths, and opportunities), evaluation of input (student preparation, curriculum of teacher performance subject, means of practice, support), process assessment (participant performance industrial work practice, teacher performance guide, performance guide business and industrial world, and product evaluation (benefits for students and benefits for business and the industrial world (Yumaroh et al., 2014).

The presence of valid and reliable internship evaluation instruments should support the importance of the internship program for vocational school students. The internship program needs a standard instrument for evaluating the success of the internship program (Rohman et al., 2020) or improving the implementation of the internship program. The existing evaluation tool used by schools in internship programs was the evaluation based on school; any school evaluates an internship program with focus group discussion. The instrument developed in this study provides convenience for schools in evaluating internship programs, especially in accounting expertise, because it consists of three aspects (preparation, implementation, and results).

This research discusses the development of the evaluation instrument of the internship competence of vocational schools in accounting expertise. The internship evaluation instruments can be used by vocational schools in conducting internship programs. The construction development of evaluation instruments needs to test the reliability and validity of the construction. Testing of evaluation instruments is carried out to determine the qualification and consistency of the evaluation instrument to evaluate the internship program of vocational school accounting students.

An instrument's validity and reliability can be tested using factor analysis. Factor analysis is a set of methods used to test how the underlying structure affects the response to several measured variables (A.Pett et al., 2003). The method used in this study is confirmatory

factor analysis (CFA). CFA tests whether a particular set of constructions affects responses predictably (A.Pett et al., 2003). CFA is used to validate factor loads and measurements involved in this study. The researchers use this approach to test the proposed theory (CFA is a form of structural equation modeling and differs from EFA, which has assumptions and expectations based on previous theories about the number of factors and which theory or model of factors is most suitable (Jian et al., 2020). This research aims to develop an evaluation instrument for vocational schools' competence in accounting expertise. The evaluation instruments were constructed based on literature reviews that tested their validity and reliability.

Literature review

Germany was an early founder of dual-system education. Germany had an early vocational education and training system called the "dual system", which many countries worldwide admired and wanted to adopt. However, to transfer this model to other countries, it is not enough to simply copy it. Rather, the transfer should reflect the conditions in the country adopting this system and be adapted according to its unique social, cultural, and economic goals. In addition, while the dual system can serve as an example, no other country can implement the dual system in its entirety or its components individually. Instead, five constitutive elements must be used to transfer the German model's dual spirit properly. The article also discusses some considerations in the transfer process and mentions some German projects that transferred the system to Asian countries (Hummelsheim & Baur, 2014). (2022) have researched the quality of Indonesian vocational-technical education graduates, which is low and does not meet industry requirements and competencies. This happens because the talents or skills the sector requires differ from those currently available. Therefore, developing countries need a practical, dual-system vocational education system. Indonesia has recently attempted to examine its Dual Vocational Education and Training system in more detail. The results show that direct or implementation transfers currently have many challenges and difficulties involving both the public and public-private sectors. On the other hand, for a successful vocational transfer, close cooperation between all stakeholders and a deeper understanding and knowledge of the contextual conditions in the target country is essential.

The implementation of the internship consists of planning and organizing, implementation, and evaluation (Astuti et al., 2023). Planning and organizing include the MoU between vocational school and industry and implementing the Training of Trainers (ToT). The implementation of the internship includes the learning process, curriculum implementation, teaching staff preparation, learning system implementation, and competency tests. The evaluation includes reporting learning progress, obtaining a certificate of competency from the industry, monitoring and evaluation, and periodic evaluations between schools and the industry. The impact of internship programs can increase labor participation, prevent early school leaving (Alegre et al., 2015), relate to continue profession, practical courses, and curriculum together in enterprises (Turan, 2022), and assist work readiness construct (Kapareliotis et al., 2019). The development of internship instruments in vocational schools in accounting expertise has not been carried out, so it is necessary to develop an evaluation instrument for accounting internship programs.

Methodology

This development research aims to prove the construction validity and estimate the

reliability of the evaluation instrument internship competence of accounting expertise in vocational school, based on the literature review, then compile the kits of instruments to continue with the development of the components of evaluation instruments. Three experts subsequently validated the development of this instrument. This instrument is a quantitative instrument. The instrument was a questionnaire filled out by the students totaling 31 items, including 13 items from internship preparation, 14 from internship implementation, and four from the internship results. Then, out of 31 items, seven outlier items and 1 item have an anti-image correlation ≤ 0.5. Thus, the total of the evaluation instruments of the internship program is 23 items consisting of 9 items from the preparation of internship (A1, A2, A3, A5, A7, A9, A10, A12, A13), ten items from internship implementation (B1, B3, B6, B7, B8, B9, B11, B12, B13, B14), and four items from internship results (C1, C2, C3, C4). Purposive sampling techniques were used to conduct the evaluation instruments' trial test at three vocational schools in the Sleman District. The purposive sampling criteria were vocational schools ranked one to three in Sleman District. One hundred eighty students responded to the survey for this evaluation tool. A student survey is used to construct the evaluation instrument for the internship programs. The questionnaire consisted of three evaluation components, including (a) the preparation of the internship(A), (b) the implementation of the internship(B), and (3) the results of the internship(C).

CFA analysis was used to estimate the reliability and validity of the construction of the evaluation instrument using CFA analysis. CFA is a statistical and qualitative process for estimating the reliability of elements, construction reliability, content validity, display validity, quantitative measurement of discriminatory and convergent validity, and goodness of fit (Hair et al., 2020). CFA is used when researchers test the hypothesis that there is a proposed theoretical relationship between the observed variable and its underlying latent structure, as well as to confirm the measurement properties of a set of variables (indicators) to measure a defined and operatively defined latent construction (Hair et al., 2020). The data was analyzed by using R software. Criteria for the conformity index of the CFA model based on test statistic, df, p-value, root mean squared error of approximation (RMSEA), Tucker Lewis Index (TLI), and comparative fit index (CFI). Model compatibility criteria using Gana and Broc criteria (Gana & Broc, 2018), an RMSEA value less than equal to 0.05 indicates a very good fit and an RMEA value smaller than 0.06 and less than equivalent to 0.08 is a good fit. A TLI value greater than 0.9 and less than equal to 0.94 indicates a good fit and a TLI value greater than 0.95 is a very good fit. A CFI value of more than 0.9 and less than equal to 0.94 indicates a good fit and a CFI rating of more than 0.95 indicates a very good fit.

Results

This study tracked variables such as internship preparation, implementation, and actual internship results. This internship evaluation tool uses CFA for validity testing and calculation of construction reliability. The CFA results demonstrate how the measurement construction model and the data from the internship evaluation matched. Additionally, the content validity of these evaluation tools was examined.

Confirmatory factor analysis of the correlation model

Data analysis of internship evaluation instruments uses CFA with a factor correlation model internship assessment. The results of the CFA analysis were then tested using the model matching criteria. Based on the results, the analysis uses the Maximum Likelihood

estimator, with the sum of model parameters 63 and the number of observations 180. The p-value value (chi-square) of the user test model is 0,000, while if the baseline test of the model, the p-value is 0.000. The p-value is not significant due to the small number of respondents. Table 1 shows the results of CFA analysis before and after the modification based on the goodness of fit test criteria. Other CFA results include RMSEA, TLI, and CFI.

Table 1: Goodness fit of test criteria value Criteria.

Criteria	Cut off value	Before modification	Status	After modification	Status
X^2	$\leq 2df$	680.732	Not Fit	346.719	Fit
RMSEA	≤ 0.05	0.105	Not Fit	0.059	Fit
TAG	≥ 0.90	0.691	Not Fit	0.903	Fit
CFI	≥ 0.90	0.723	Not Fit	0.918	Fit

The goodness of fit test can be seen with RMSEA, TLI, and CFI values. RMSEA = 0.059, meaning > 0.05 so, indicating a good fit of the model. The TLI value is 0.903, meaning < 0.95, indicating a good fit model. The CFI value is 0.918, meaning <0.95, indicating a good fit model. Based on the results, the model of evaluation instrument internship competence skills Accounting SMK developed fit.

The latent variables of this study include variable P (internship preparation), variable I (internship implementation), and variable R (internship results). Figure 1 and Table 2 report information about each latent variable's loading factors. The variable P, measured by the element A1, has a load factor value of 0.378. The variable P measured by element A2 has a load factor value of 0.301. The variable P measured by element A3 has a load factor value of 0.364, and so on. The highest load factor value is the variable P measured by element A9, which is 0.805. Variable I, measured by element B1, has a load factor value of 0.265. Variable I, measured by element B3, has a load factor value of 0.376. Variable I, measured by element B6, has a load factor value of 0.482. The highest load factor value is variable I, measured by element B7, which is 0.837. The variable R measured by the element C1 has a load factor value of 0.265. The variable R measured by the element C2 has a load factor value of 0.632. The highest load factor value is the variable R measured by the element C3, which is 0.761. The variable R measured by the element C4 has a load factor value of 0.606. The result of the modification shows that variable element A13 also measures me with a load factor value of 0.311.

The results of instruments that significantly influence the observation variables are seen from the standardized factor load (SFL) values in Table 2. A particle or indicator with good structural validity and a good latent variable if the SFL value is ≥ 0.30 [19]. Based on Table 2, all the seeds meet the requirements; only B1, B9, B11, and B14 have values < 0.30. However, the SFL value of the four grains is more than 0.25 (close to 0.30).

Table 2: Standardized factor load (SFL)

Components	Items	SFL
Preparation of internship	A1	0.378
	A2	0.301
	A3	0.364
	A5	0.545
	A7	0.562
	A9	0.805
	A10	0.764
	A12	0.707
	A13	0.390
Implementation of internship	B1	*
	B3	0.376
	B6	0.482
	B7	0.837
	B8	0.803
	B9	*
	B11	*
	B12	0.779
	B13	0.558
	B14	*
	A13	0.311
Results of internship	C1	0.709
	C2	0.632
	C3	0.761
	C4	0.606

* SFL not meeting the criteria

Information related to covariance or correlation between the seeds and other seeds is shown in Figure 1. The highest covariant value is A10 with A12, which is 0.832. The covariance value of B9 with B11 is 0.586. The covariance value of A7 with A13 is 0.423. The variation value of variable P with variable I is 0.192. The value of variable P for variable R is 0.146. The covariance value of variable I with variable R is -0.021, and so on.

Figure 1 also provides information about variance values. The variance value of A1 is 0.857. The variance value of A2 is 0.910. The variance value of A3 is 0.868. The variance value of A5 is 0.703. The variance value of A7 is 0.684. The variance value of A9 is 0.351, and so on. The highest variance value is B1, which is 0.930. The lowest variance value is B7, which is 0.299. The variables P, I, and R variance values are 1,000. The value of R-square, or (the size of each element's contribution) is shown in Table 3. The highest contribution value is B7, which is 0.701. The lowest contribution value is B1, which is 0.070.

Table 3. R-square

Items	Estimate R-square
A1	0.143
A2	0.090
A3	0.132
A5	0.297
A7	0.316
A9	0.649
A10	0.584
A12	0.500
A13	0.296

Items	Estimate R-square
B1	0.070
B3	0.141
B6	0.232
B7	0.701
B8	0.644
B9	0.073
B11	0.082
B12	0.607
B13	0.311
B14	0.072
C1	0.503
C2	0.400
C3	0.579
C4	0.367

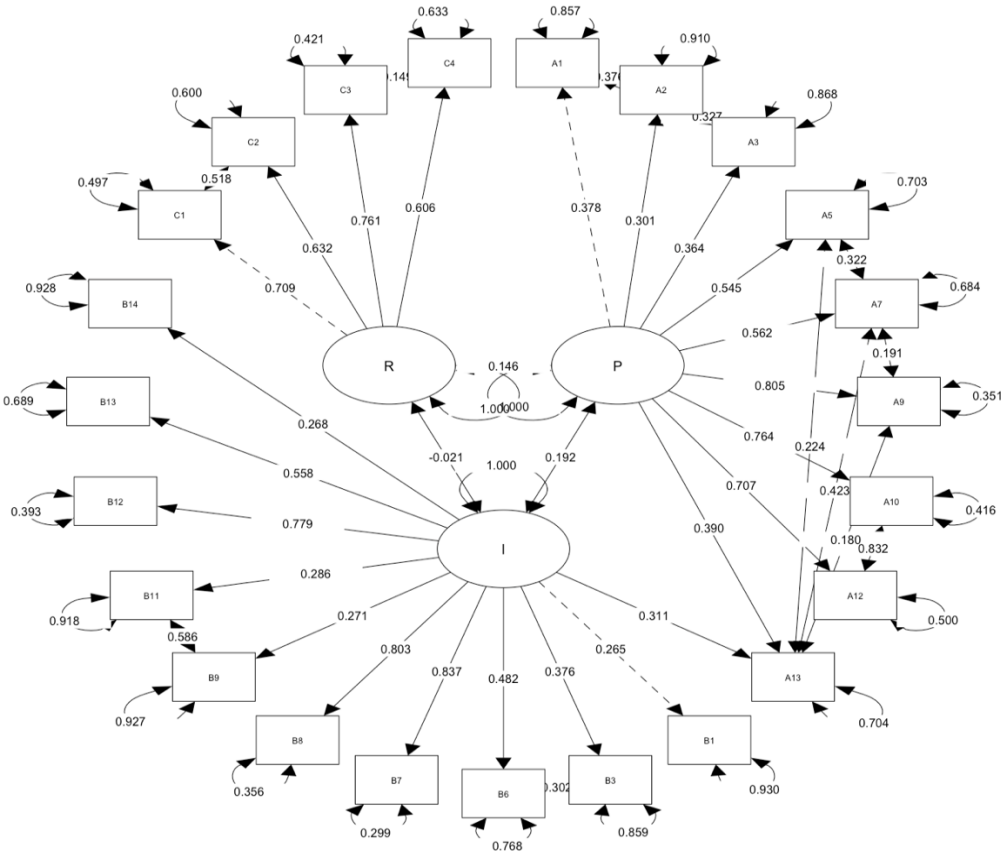

Figure 1. Standardized solution diagram of CFA results

The analysis findings utilizing the CFA factor correlation model are displayed in Figure 1. To modify the outcomes of this CFA analysis, we connected the elements and performed a correlation between items A10-A12, B9-B11, A7-A13, A5-A7, A1-A2, A6-A9, A2-A3, C1-C2, C3-C4, A9-A13 and the component I in the A13 scale.

Construct validity and composite reliability.

The theoretical construction, also known as the validity of the construction, must be consistent with the instrument's development. In order to refer to the meaning of

measurement, construction validation is the process of obtaining evidence (Hair et al., 2020). Once the suitability requirements of the measurement model are met, the developed model needs to be tested for validity and reliability estimates before proceeding with the structural equation model (Hair et al., 2020). The information in Table 1 shows that based on the index of conformity of the measurement model, all index criteria are met in the validation of the construction of the evaluation instrument of the internship program. Therefore, the validity of the construction of the evaluation instrument of the internship program is valid for evaluating the accounting student who follows the industrial work practice.

The average variance extract (AVE) is 0.35. According to Shrestha (Shrestha, 2021), the convergent validity size must have an AVE value greater than 0.50 (Nasution et al., 2020)**.** The result of an AVE value shows that each observed variable has a value less than 0.50. This means that the latent structural variance for each indicator element is lower than the error variant, so the internship program evaluation instrument element is less valid for measuring each observed variable. Therefore, the internship program evaluation instrument has less convergence validity in measuring each variable observed on the internship program evaluation instrument.

The next analysis concerns the validity of discrimination. Discriminatory validity is sometimes presented as the property of construction and as a property of the size built from that size (Rönkkö & Cho, 2022). A discriminatory validity value is used to analyze the relationship between latent variables (Henseler et al., 2015). Discriminatory validity can be defined as the correlation between two constructions. If the correlation value of the two structures is less than 0.85, the instrument's construction meets the discriminatory validity criteria (Noor et al., 2015). The CFA model correlation of the internship program evaluation instrument obtained correlations between the warning variables P-I, P-R, and I-R, respectively, at 0.192, 0.146, and -0.021. The study's results showed that the evaluation instrument of the internship program meets discriminatory validity with a correlation value of less than 0.85. Thus, the evaluation instrument of the internship program has a different warning variable structure than the other warning variables.

Testing continued with the reliability estimate of the evaluation instrument of the internship in the accounting program. An instrument can be said to be reliable if it measures the same ability repeatedly with consistent results. The reliability of compounds is discussed in this study. Composite reliability can also be called internal consistency (Hair et al., 2017), a combination of latent construction reliability that underpins the measurement scale (Ridwan et al., 2023). Composite feasibility can be used as a measurement tool but is not recommended for selecting particles in a structural equation model (Bacon et al., 1995). The reliability criterion of this composite is greater than 0.7 (Hair et al., 2020). The composite reliability of the composite rho value of this study is 0.92. The estimated reliability of this composite is greater than 0.7. This indicates that the evaluation instrument of the internship competence of the vocational school in accounting expertise provides consistent results.

Content validity

Test the validity of the content of this fieldwork practice evaluation instrument using three indicators: preparation for fieldwork practice, implementation of fieldwork practice, and the results of fieldwork practice. The proof of validity uses content validity based on three

experts analyzed using the Aiken formula. Based on the results of proving content validity in Table 3, it shows that the 23 evaluation instruments are feasible or valid for assessing the SMK internship program for accounting expertise competency.

Table 3. Content Validity Results

Items	V	Status
A1	0.8333333	Valid
A2	0.8333333	Valid
A3	0.7500000	Valid
A5	0.8333333	Valid
A7	1.0000000	Valid
A9	1.0000000	Valid
A10	0.7500000	Valid
A12	0.7500000	Valid
A13	0.7500000	Valid
B1	0.8333333	Valid
B3	0.8333333	Valid
B6	0.7500000	Valid
B7	0.7500000	Valid
B8	0.8333333	Valid
B9	0.7500000	Valid
B11	1.0000000	Valid
B12	0.8333333	Valid
B13	0.8333333	Valid
B14	0.9166667	Valid
C1	0.8333333	Valid
C2	0.9166667	Valid
C3	0.7500000	Valid
C4	0.9166667	Valid

Discussion

Previous research related to the evaluation of internship implementation has been conducted. Research on the evaluation of internship activities has been done by Iwa Kuntadi (2023). The results explain the need for a CIPPO evaluation model with context, input, and product aspects. The outcome aspect is necessary for improving the quality of the cooperation program between vocational schools and industry from the vocational school perspective in Western Java. This research is limited to the analysis of CIPPO evaluation needs. Waryono et al. (2023) researched to analyze the evaluation results of implementing training and distribution in IM Japan. The findings showed that context and input aspects were evaluated well, with a percentage value of 22% each; process aspects were evaluated excellently, with percentages of 49%; and product aspects need to be enhanced in the alumni assembly to develop an alumni network or participant training and dissemination related to workforce needs. This study has not discussed the development of its evaluation instruments.

Setiyawan and Kurniawan (2021) researched the usefulness of task-learning information systems and student workforce monitoring-based SMS gateways with Raspberry Pi. The tool used in the testing is the Computer Usability Satisfaction Questionnaire: Psychometric Assessment and Instructions for Use, developed by IBM for software usability measurement standards. The tool for testing usability aspects with the Computer System Usability Questionnaires (CSUQ) questionnaire developed by IBM uses the Likert scale as

its measurement scale. The study results were that the usability aspect had a percentage of 85% or a high-quality scale, and based on alpha calculations, Cronbach had a calculation result of 0.851 or the category “good”. The SMS Gateway system with the Raspberry Pi for internship assignment and monitoring meets the usability aspects. However, this study has not discussed the correlation testing of the instrument's development model.

Research on developing a competence-based performance assessment model for sharing family welfare education (Jubaedah et al., 2017). The result of this research is the planning, instrumentation, and implementation of assessments in developing assessment models based on the competence of work performance in disseminating family welfare education. Assessment planning includes the components of objectives, work performance, and assessment methods. The assessment instrument uses a work demonstration test as an assessment section. Execution of assessments includes the stages of preparation, collection, adjustment, dismantling, and moderation. Based on the validation model through expert judgment, the results of the assessment model developed are suitable for implementation in the performance assessment of the sharing of family welfare education. This development research has not dealt with the matching testing of the developed instrument model. Experts validate the validity test; there are no validity tests of construction or other validity and reliability tests.

Azman et al. (2020) conducted the development of program evaluation instruments used to evaluate the internship program of computer and networking engineering competence in vocational schools, including input components (input), among others: students, productive teachers, curriculum, and facilities, process components, among other things: student preparedness, accompanying teacher performance, business, and industrial world accompaniment performance, the output of the internship program utility for students, as well as business and industrial world. The results of this study were that input components were 98.9%, process 96.6%, and product 96.7%, and the instruments developed were already valid and fit very well. Student questionnaire reliability coefficient: 0.963; productive teacher: 0.981; accompanying teacher: 0.942; associate world of enterprise and industry: 0.961. However, the study did not address the model-matching testing of the instruments developed. So, the conclusion is not strong because considering the development of instruments is only a test of validity and reliability.

Negara & Hidayati (2021) conducted research aimed at developing tools for assessing the performance of internships with the Work-Based Learning (WBL) approach and Learning Innovation Skill (LIS) concerning the Framework of Partnership 21st Century Skills. The study carries out three validity tests: expert validity, content validity, and criteria validity. The results of the content validity analysis using Aiken's V formula obtained for the WBL variable have a 50% high validity percentage and 50% medium validity. In comparison, the LIS variable has 89.74% for high validation and 10.26% for moderate validity. In this test, the number of items for WBL variables was 32, and those for LIS variables were 39. The validity test of the criteria is analyzed using the Product Moment Correlation. For the WBL variable, two elements are declared valid, so both elements are invalid, while the LIS variable for all elements is valid. Furthermore, reliability testing uses Cronbach's Alpha, where the WBL variable has a reliability of 0.942, and the LIS variable has a reliability of 1,000, so it can be concluded that the instrument developed is declared valid and reliable to assess student WBL and LIS activities. However, this study has not tested the matching of the models of the instruments developed.

Research conducted by To (2017) on the program evaluation of the internship program operated by SiemensEnergy, Inc. in Charlotte, North Carolina, based on the Stufflebeam CIPP model. Four questions were developed to align with the CIPP model: (a) context: how the program objectives are tailored to the needs of Siemens and the interns; (b) introduction: what characteristics help traders complete their program? (c) process: Have traders been successfully trained? (d) Product: What results in meeting the program's strategic plan? The research design methodology follows several paths. The evaluation results of this program are effective internship programs and the achievement of the objectives. Researchers have concluded three recommendations, including providing more support to interns, ensuring interns fully understand the curriculum requirements and expectations for their associate degree, and developing a strategic plan with a statement of mission and vision. The research focuses on evaluating internship programs but has not discussed the development of evaluation instruments in detail. Based on various relevant research results, this research contributes to developing the evaluation instrument for the vocational school, especially the accounting program. The development of evaluation instruments has been carried out through CFA analysis and model matching testing, resulting in 23 valid instruments that can be used to evaluate the internship program evaluation in vocational schools. The respondent of this instrument was a student. For further research, researchers can develop an internship evaluation instrument that student mentors or supervisors can fill out at the internship location.

Conclusions

The study uses the CFA with the correlation factor model to prove the validity and reliability of the evaluation instrument construction of the vocational school. Testing of the research data has been modified to obtain a measurement model that matches the evaluation capability data. The evaluation component of the vocational school internship program in accounting expertise consists of internship preparation, internship implementation, and internship results. The content validity test, construction validity, and construction reliability estimate result in 23 valid instruments that can be used to evaluate the internship program evaluation of accounting field practice in a vocational school. This tool is effective in evaluating accounting internships. Recommendations for further research include the development of website-based internship program evaluation instruments so that the instruments are more easily accessible to respondents and more effective in data collection.

References

Alegre, M. À., Casado, D., Sanz, J., & Todeschini, F. A. (2015). The impact of training-intensive labor market policies on labor and educational needs prospects: Evidence from Catalonia (Spain). *Educational Research*, *57*(2), 151–167. Scopus. https://doi.org/10.1080/00131881.2015.1030852

Anjum, S. (2020). Impact of internship programs on professional and personal development of business students: A case study from Pakistan. *Future Business Journal*, *6*(1), 2. https://doi.org/10.1186/s43093-019-0007-3

A.Pett, M., R.Lackey, N., & J.Sullivan, J. (2003). *Making Sense of Factor Analysis*. SAGE Publications, Inc. https://doi.org/10.4135/9781412984898

Astuti, A. N. F., Samsudi, S., Sudana, I., Made., & Hadromi, H. (2023). *Exploring Management Industrial Class at the Vocational High School in Indonesia | Journal of Educational and Social Research*. https://www.richtmann.org/journal/index.php/jesr/article/view/13345

Azman, A., Giatman, M., Ambiyar, & Syahril. (2020). Pengembangan Instrumen Evaluasi Program Prakerin Kompetensi Keahlian TKJ di SMK. *Cakrawala: Jurnal Pendidikan*, 64–71. https://doi.org/10.24905/cakrawala.v14i1.1512

Bacon, D. R., Sauer, P. L., & Young, M. (1995). Composite Reliability in Structural Equations Modeling. *Educational and Psychological Measurement*, *55*(3), 394–406. https://doi.org/10.1177/0013164495055003003

Deissinger, T. (2010). Dual System. In *Publ. In: International Encyclopedia of Education. Vol. 8 / Penelope Peterson ... (Eds.). 3. Ed. Oxford: Elsevier, 2010,* pp. 448-454.

Deissinger, T., & Gonon, P. (2021). *The German system of vocational education and training: Challenges of gender, academisation and the integration of low-achieving youth—Simone R Haasler, 2020*. https://journals.sagepub.com/doi/10.1177/1024258919898115

Disas, E. P. (2018). Link and Match sebagai Kebijakan Pendidikan Kejuruan. *Jurnal Penelitian Pendidikan*, *18*(2), Article 2. https://doi.org/10.17509/jpp.v18i2.12965

Gana, K., & Broc, G. (2018). *Structural Equation Modeling with lavaan*. https://doi.org/10.1002/9781119579038

Haasler, S. R. (2020). *The German system of vocational education and training: Challenges of gender, academisation and the integration of low-achieving youth*. https://www.ijopr.com/home/author-guidelines

Hair, J. F., Howard, M. C., & Nitzl, C. (2020). Assessing measurement model quality in PLS-SEM using confirmatory composite analysis. *Journal of Business Research*, *109*, 101–110. https://doi.org/10.1016/j.jbusres.2019.11.069

Hair, J. F., Matthews, L., Matthews, R. L., & Sarstedt, M. (2017). *PLS-SEM or CB-SEM: Updated guidelines on which method to use*. https://www.semanticscholar.org/paper/PLS-SEM-or-CB-SEM%3A-updated-guidelines-on-which-to-Hair-Matthews/0a4787bf5c0e58fa43ff45c9874140b0d421002c

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Hummelsheim, S., & Baur, M. (2014). The German dual system of initial vocational education and training and its potential for transfer to Asia. In *Prospects* (Vol. 44, Issue 2, pp. 279–296). Kluwer Academic Publishers. https://doi.org/10.1007/s11125-014-9311-4

Ismail, R. (2022). Partisipasi Dunia Usaha Dunia Industri Dalam Peningkatan Mutu Lulusan Sekolah Menengah Kejuruan Se Kota Gorontalo. *Jurnal Informatika dan Teknologi Pendidikan*, *2*(1), 23–27. https://doi.org/10.25008/jitp.v2i1.22

Iwa Kuntadi, E. S. (2023). A needs analysis of Cippo evaluation model components as a quality escalation of the cooperation program with industries: vocational school perceptions in west Java, Indonesia. *Journal of Southwest Jiaotong University*, *58*(1), Article 1. http://jsju.org/index.php/journal/article/view/1509

Jian, O. Z., Yin, K. Y., & Awang, M. (2020). Developing And Validating the Measurement Model for Employee Engagement Construct Using Confirmatory Factor Analysis. *International Journal of Academic Research in Business and Social Sciences*, *10*(8), 924–941.

Jubaedah, Y., Rohaeni, N., & Rinekasari, N. R. (2017). Development of Competency-Based Assessment Model on Job Performance in Family Welfare Education Apprenticeship. *INVOTEC*, *13*(1), Article 1. https://doi.org/10.17509/invotec.v13i1.6262

Kapareliotis, I., Voutsina, K., & Patsiotis, A. (2019). Internship and employability prospects: Assessing student's work readiness. *Higher Education, Skills and Work-Based Learning*, *9*(4), 538–549. https://doi.org/10.1108/HESWBL-08-2018-0086

Maulina, M., & Yoenanto, N. H. (2022). Optimalisasi link and match sebagai upaya relevansi SMK dengan dunia usaha dan dunia industri (DUDI). *Jurnal Akuntabilitas Manajemen Pendidikan*, *10*(1), Article 1. https://doi.org/10.21831/jamp.v10i1.48008

Nasution, M. I., Fahmi, M., Jufrizen, Muslih, & Prayogi, M. A. (2020). The Quality of Small and Medium Enterprises Performance Using the Structural Equation Model-Part Least Square (SEM-PLS). *Journal of Physics: Conference Series*, *1477*(5), 052052. https://doi.org/10.1088/1742-6596/1477/5/052052

Negara, I. M., & Hidayati, N. (2021). The Development of Assessment Instrument to Assess Implementation of Work-Based Learning and Learning Innovation Skills. *Jurnal Iqra' : Kajian Ilmu Pendidikan*, *6*(1), Article 1. https://doi.org/10.25217/ji.v6i1.932

Noor, N. M., Aziz, A. A., Mostapa, M. R., & Awang, Z. (2015). Validation of the Malay Version of the Inventory of Functional Status after Childbirth Questionnaire. *BioMed Research International*, *2015*, e972728. https://doi.org/10.1155/2015/972728

Ramadhani, M. A., & Rahayu, E. (2020). Competency improvement through internship: An evaluation of corporate social responsibility program in vocational school. *International Journal of Evaluation and Research in Education (IJERE)*, *9*(3), Article 3. https://doi.org/10.11591/ijere.v9i3.20571

Regulation of the Minister of Education and Culture of the Republic of Indonesia No. 50 of 2020. (n.d.).

Remington, T. F. (2018). Public-private partnerships in TVET: adapting the dual system in the United States. In *Journal of Vocational Education and Training* (Vol. 70, Issue 4, pp. 497–523). Routledge. https://doi.org/10.1080/13636820.2018.1450776

Ridwan, M. R., Hadi, S., Jailani, J., & Retnawati, H. (2023). The instrument is developed to measure the verbal ability of prospective high school students. *International Journal of Evaluation and Research in Education (IJERE)*, *12*(1), Article 1. https://doi.org/10.11591/ijere.v12i1.22736

Rohman, S., Samsudi, & Sutopo, Y. (2020). *Need Analysis Development Instruments Evaluation of Vocational Industrial Work Practices Based on Competence Fitness*. 603–607. https://doi.org/10.2991/assehr.k.200620.123

Rönkkö, Mi., & Cho, E. (2022). *An Updated Guideline for Assessing Discriminant Validity*. https://doi.org/10.1177/1094428120968614

Saputra, B. D., & Sukirno, S. (2020). Kesiapan kerja siswa program akuntansi pada sekolah menengah kejuruan. *Jurnal Kependidikan Penelitian Inovasi Pembelajaran*, *4*(1), Article 1. https://doi.org/10.21831/jk.v4i1.24651

Setiyawan, A., & Kurniawan, A. (2021). The Effect of Pandemic Covid-19 on Internship Activity of Mojokerto Vocational High Schools. *Journal of Vocational Education Studies*, *4*(1), 125–130. https://doi.org/10.12928/joves.v4i1.4080

Shrestha, N. (2021). Factor Analysis as a Tool for Survey Analysis. *American Journal of Applied Mathematics and Statistics*, *9*(1), Article 1. https://doi.org/10.12691/ajams-9-1-2

Sudira, P. (2019). The Role Of Vocational Education In The Era Of Industrial Automation. *Journal of Physics: Conference Series*, *1273*(1), 012058. https://doi.org/10.1088/1742-6596/1273/1/012058

Suharno, Pambudi, N. A., & Harjanto, B. (2020). Vocational education in Indonesia: History, development, opportunities, and challenges. *Children and Youth Services Review*, *115*, 105092. https://doi.org/10.1016/j.childyouth.2020.105092

To, O. C. (2017). *A Program Evaluation of an Apprenticeship Program using Stufflebeam's CIPP Model*.

Turan, S. N. (2022). Research on the effect of the perception of the school and the sector on students' intention to continue. *Milli Egitim*, *51*(235), 2533–2558. Scopus. https://doi.org/10.37669/milliegitim.928355

Waryono, E., Ahmad, M., & Rahmawati, D. (2023). Evaluating Training and Apprenticeship as Non-Formal Education at International Manpower Development Japan. *Jurnal Paedagogy*, *10*(2), 320–329. https://doi.org/10.33394/jp.v10i2.6994

Wibowo, R. A., Myau-Lyau, N., Christy, N. N. A., & Albert. (2022). The Challenges for Indonesia to Integrate Dual Vocational Education and Training System. In *Journal of Technical Education and Training* (Vol. 14, Issue 2 Special Issue, pp. 79–90). Penerbit UTHM. https://doi.org/10.30880/jtet.2022.14.02.008

Yumaroh, Y., Lestari, W., & Masrukan, M. (2014). Pengembangan instrumen evaluasi program praktik kerja industri kompetensi keahlian teknik audio video smk. *Journal of Research and Educational Research Evaluation*, *3*(2), Article 2. https://journal.unnes.ac.id/sju/index.php/jere/article/view/4398