European Journal of Educational Research
Volume 10, Issue 3, 1529 - 1540.
ISSN: 2165-8714
https://www.eu-jer.com/

Assessments on Vocational Knowledge and Skills: A Content Validity Analysis

Mazlili Suhaini*
Universiti Teknologi Malaysia, MALAYSIA

Adnan Ahmad
Universiti Teknologi Malaysia, MALAYSIA

Normila Mohd Bohari
Universiti Teknologi Malaysia, MALAYSIA

Received: February 6, 2021 ▪ Revised: April 8, 2021 ▪ Accepted: June 27, 2021

Abstract: While previous studies on assessments focused on measurement of principles and constructs, existing attention is given to content validation involving vocational education and skills. The emphasis on content validation has prompted a holistic perspective of teaching and learning to demand alternative research approaches. Using evidence of content validity based on expert judgment and assessment items, this paper argues that vocational knowledge and skills could be determined via constructs and construct functions. Content validity analysis was studied in two major phases, namely, through an assessment of an electrical technology course at a Malaysian higher education institution and expert panels' examination of items. It was found that to increase confidence in providing reliable instruments for future empirical studies, a careful process in item development and content validity analysis was considered important. Therefore, applying these findings on item analysis and expert panels to reflect content validity can enhance the validity of assessment items.

Keywords: *Assessments, content validity analysis, I-CVI, modified kappa, vocational knowledge.*

To cite this article: Suhaini, M., Ahmad, A., & Mohd Bohari, N. (2021). Assessments on vocational knowledge and skills: A content validity analysis. *European Journal of Educational Research, 10*(3), 1529-1540. https://doi.org/10.12973/eu-jer.10.3.1529

Introduction

The term, assessment, which indicates its Latin origins, 'assider', means "to sit on the side" (James, 2010). In particular, the word 'assessment' gives the impression that assessments, in general, demonstrate students' thorough understanding of learning under constant supervision by teachers. The context of a general perspective of assessment highlights two important functions of assessments, namely, to map student achievement levels and satisfy educational accreditation bodies. Firstly, assessments were designed to measure the extent to which assessments reflected actual learning and real-world contexts (Ghafar, 2011; Saha, 2021). As such, building assessments involved a process of gathering, analysing and interpreting student achievement levels to meet teaching objectives (Russell & Airasian, 2011). Secondly, assessments are widely associated with governing accreditation bodies. Within a mainstream education, a single national accreditation body stipulated that measurements were derived from evaluation processes that functioned to assign numerical values to attributes, characteristics, and achievement of learning outcomes (Agency, 2013).

Conducting an assessment could really assist an educator in determining the level of understanding of students' knowledge after learning an ability (Ulum et al., 2021). As a consequence, in order to improve students' thinking abilities, assessment instruments should be properly designed and suited to each student's level of cognitive abilities. As is well known, assessment should be carried out not only at the end of the lesson but also while the teaching and learning process is in progress.

Assessments are generally understood in two broad features, namely, functions and contexts. The Malaysian Institute of Teacher Education described that assessments functioned to provide an overview of student learning performance (Malaysian Institute of Teacher Education, 2019). An overview of integrated student learning performance should begin with formative assessments (Black & Wiliam, 2018). As such, formative assessments, which are usually referred to as classroom assessments, could promote student learning in classrooms based on immediate teachers' feedback. Secondly, assessments largely depend on classroom learning purposes. For instance, assessment components could measure one's ability to manipulate limbs through practical, hands-on components (Hamid et al., 2012) and present

* **Corresponding author:**
Mazlili Suhaini, Universiti Teknologi Malaysia, School of Education, Malaysia. ✉ mazlili.2385@gmail.com

knowledge and theoretical foundation (Agarwal, 2019). Therefore, assessment functions and contexts remain fundamental in the understanding of the relationship between classroom learning and how learning is measured.

Therefore, collecting feedback from students to be analysed is key to interpret and determine student achievement in teaching and facilitation processes. Thus, a formative measurement of student achievement, that accumulates feedback from students, can continuously support the measurement of student achievement summatively or as a whole. By focusing on student feedback, the results of the analysis are then employed to assess student achievement outcomes, benefiting students, teachers and school administrators.

Literature Review

Technical and Vocational Education (TVET)

One of the strategies developed by the Malaysian government to promote technical and vocational education training (TVET) could be seen through the implementation of the 11th Malaysia Plan (RMK-11). As highlighted in RMK-11, TVET programmes were designed so workers could learn prospective skills that all sectors and industries demanded through Malaysia's creation of a 60% increase in TVET-related opportunities. As opposed to only 25% of highly-skilled workforce employed (Malaysian Ministry of Education, 2013), RMK-11 was instrumental in organising the existing highly-skilled workforce in observance of Malaysia's initiative to accomplish the status of a developed country. Therefore, providing TVET educational opportunities is a key predictor to grow highly-skilled workers.

Local and global alignment of TVET had already taken place. Within Malaysia’s Education Blueprint 2015-2025, the emphasis on TVET was highlighted as one initiative that was planned for the future of Malaysia. As outlined in Malaysia’s Education Blueprint, the primary importance of TVET could be seen in Malaysia’s emphasis on vocational education and skill development that began in 1906. Nevertheless, TVET is realised as an aspect of educational process and general education involving learning in technology and related sciences, including training in practical skills, attitudes, understanding, and knowledge of employment at various sectors of economy and social life, drawn based on the United Nations Organisation for Education, Science and Culture (UNESCO)’s definition (Malaysian Ministry of Education, 2015).

Malaysian government supports have now taken the efforts to expand access to TVET education to another dimension in its efforts to grow highly-skilled students (Ekonomi, 2016). For instance, through the establishment of *Vocational College* (VC) institutions, opportunities to boost excellence in skills were evidenced. In particular, the focus of VC on the product of revamped Vocational secondary schools was needed to accommodate changing and challenging labour market trends and higher vocational education degrees (Jalal et al., 2017; Omar et al., 2018). The curriculum of VC that was compiled was seen as an integration of the principles of job competency set out by the organisation of workplace system and how work-based skill components met and streamlined the requirements of the Malaysian Qualifications Framework scale (Malaysian Ministry of Education, 2012). By accommodating the needs of workplace labour and national quality control, the organisation of curricula by emphasising learning vocational training through different approaches could demonstrate students’ abilities.

Assessments in TVET

Technically, TVET education qualifications exhibit a variation in terms of content. Firstly, in TVET education, a certain number of academic and vocational subject hours are provided as a means of training TVET majors to master the relevant knowledge, skills, and attitudes in the course of their studies (McGrath, 2014). Depending on the nature of the courses, TVET education primarily focuses on theoretical knowledge of the courses such as identifying the number of components, explaining the types of components found in the system, and demonstrating the operationalisation of the systems learnt in the course.

Secondly, TVET education is also composed of competency-based assessments (CBA). Institutions or training centres generally prepare students to sit for training that combines theoretical knowledge and skills so TVET majors will be able to solve problems, for example, solving issues involving a determination of disrupted components or systems. By focusing on TVET education through a combination of knowledge and skills equivalence, students could work on fixing components or systems so that the sequence, scope, and selection of components and systems will function. As such, success in educating TVET majors is also connected to the extent of how training institutions or centres combine theories and skills. Thus, not only does TVET education involve a mastery of knowledge and practical skills, writing and hands-on practice are also important to carry out assignments that meet TVET curriculum standards.

Competency-based assessment (CBA)

By broad definitions, job competency encompassing knowledge, skills, and capacity to conduct a profession by meeting existing standards (Daniel & Hultin, 2002; Kyobe & Rugumayo, 2005; Meyer-Adams et al., 2011) is commonly accomplished through competency-based assessment (CBA). CBA emphasised the extent to which employees achieve and meet job expectations in a specific industry or career (Likisa, 2018). Moreover, by combining practical and theoretical knowledge, CBA offered a curriculum framework that distinguished the dichotomy between "knowing that" and "how to know" (Gonczi, 1996), that is streamlined with the needs, objectives, and scope in meeting workplace demands (Education, 2013). Therefore, CBA guides and regulates the transition from knowledge acquisition to knowledge application, meeting learning objectives successfully.

The overall aims of CBA heavily rely on practical applications. To a varying extent, TVET education has wider aims to train competent TVET majors for vocational preparation, granting eligibility for entry into labour market. Competence-based assessment combines the assessment of students, particularly in the field of education, and practical, hands-on activities. The Government of Western Australia, for instance, believed that CBA was connected to a process of gathering evidence and making judgments on a set of criteria. The criteria based on the student's performance demonstrated how students performed in a work environment (Council, 2009). Therefore, Gonczi (1996) defined CBA as a method that encompassed many assessment tools that can be used to evaluate student performance. Therefore, CBA is designed to test students' knowledge, skills, and attitudes.

One of the notable CBA assessment tools is portfolios. According to Mazin et al., (2020), portfolios serve as one of the additional resources to demonstrate performance in learning in a detailed, organised format. Specifically, by compiling evidence of learning in a folder, the evidence of learning is further indexed and mapped to relevant performance criteria. Thus, CBA as a learning method, emphasises diverse assessment methods used to determine whether learning that has taken place has reached the minimum threshold of knowledge, skills and attitudes necessary to perform a specific task successfully. Therefore, CBA could be understood as a form of assessment that begins with a statement of objectives, subsequently followed through with critical evidence and basic knowledge of the qualifications for the course.

In conclusion, assessments are key predictors of student learning. Vocational education is deemed successful if CBA-related assessments carefully reflect teaching approaches and specific learning outcomes. In short, CBA as an approach to demonstrate students' mastery of knowledge and skills performed in a variety of tasks is also capable to foster positive effects on student success and enable students to develop their expertise. Thus, a holistic perspective of CBA resembles a motivation for further research for better, prospective assessments, as the ensuing discussion will show.

A study on content validity analysis

Within the design of content analysis, validity played two important factors in the selection of pilot study instruments (Lynn, 1986). Firstly, measurement in vocational education was considered acceptable if content validity was measured in the contexts of knowledge, skills, attitudes, and behaviour (Cohen et al., 2007). Specifically, validation involved how items of an instrument measured valid principles and constructs. Secondly, the instruments that were built also needed to comply with the required standards and reviews by content experts (Ghafar, 1999). By focusing on content and expert panels' judgments, items that did not reflect the connection among the content, table of test specification, and learners' performance were removed with due considerations and recommendations by expert panels.

Content validity also depended on two processes, namely, development and evaluation (Lynn, 1986). Firstly, what was achieved at the level of 'development process' involved domain tracking and generating instruments (Carmines & Zeller, 2014). The context-specific definition of constructs was generally selected through domain tracking. Items from domain tracking were subsequently produced and reorganised to meet the criteria for the next stage of evaluation.

Secondly, content validity analysis involved an 'evaluation' on units of analysis. In the second stage, the instrument was subjected to reviews by expert committees on the validity of items and the overall description of the instrument. The comments which were received by a panel of experts required discretion on the part of researchers to examine the plausibility of items that best fit the desired content domain. Following this stage of validation, reliability assessment was also used in the context of Item-Content Validity Index (I-CVI) and the modified kappa (k^*) as a resource validation tool (Polit et al., 2007).

Item-level content validity index (I-CVI)

I-CVI examined an index of agreement between expert evaluators (Polit et al., 2007). By applying agreement protocol, I-CVI required evaluations and assessments from at least three independent experts concerning the suitability of a group of items (Lynn, 1986). The suggestion of Polit et al. (2007) to resolve the issue of the relationship agreement between the I-CVI values and the modified kappa index was followed. Therefore, this process removed non-compliance agreements because non-compliance agreements did not indicate a judgment on the validity of items.

I-CVI was calculated through the total number of expert agreements. The division calculation implied that I-CVI = Number of expert agreements, which indicated a level of agreement on specific items divided by the number of expert panels. The result of the calculation demonstrated the value of I-CVI that achieved an acceptable level of item suitability. The following description of I-CVI is similar to the calculation formula of Equation 1:

$$ICVI = \left[\frac{ne}{N}\right] \qquad \text{Equation 1}$$

ne = Number of approvals on the relevant object (3 or 4)

N = Number of expert panels

Expert panels are commonly selected based on an individual's expertise in a particular subject. The recommended number of expert panels is summarised in Table 1. Lynn, (1986) suggested a minimum of three experts and a maximum of ten experts. While Davis proposed that the final decision on the number of experts required for content validity be made by at least two reviewers who are experts in the content area to be measured (Davis, 1992). Polit et al. (2007), on the other hand take a full agreement approach. The judges are basically in agreement that the item is relevant. Agreement about non-relevance is not counted because it does not inform judgments about an item's content validity. The maximum number of experts has not been specified, but up to ten experts are regularly used.

Table 1. Number of experts

Number of experts	Source of recommendation
Two experts	(Davis, 1992)
At least three to ten experts	(Lynn, 1986)
At least three experts	(Polit et al., 2007)

However, in this study, researchers are taking the approach of Polit et al. (2007). To ensure the content's validity, items in expert panels' judgment were evaluated using I-CVI. Following Polit et al. (2007), subsequent data were analysed to obtain the value of k*. Table 2 shows the criteria for evaluation of I-CVI values concerning reasonable items to be retained, reviewed, and eliminated.

Table 2. I-CVI value

Item	Criteria	Description
Item-level content validity index (I-CVI)	<0.70	To be eliminated
	0.70 - 0.79	To be reviewed
Polit et al. (2007)	>=0.80	To be retained

According to Polit et al. (2007), 'to be retained' refers to scores greater than 0.80, 'to be reviewed' ranges between 0.70 and 0.79, and 'to be eliminated' is for scores less than 0.70. The items that were eliminated lacked accuracy in measuring certain constructs and were unsuitable for designing good test instruments.

Modified Kappa analysis

Once all I-CVI values were obtained, k* value analysis could be conducted. The k* statistics, generally known as modified kappa, is a special agreement index in which the agreement between experts assessing only reasonable items is considered. Therefore, I-CVI and Pc values in this study were determined by k* statistics. Extraneous items as decided by experts were not considered. While the calculation formula for Pc is given in Equation 2, the formula for evaluating the value for k* is provided in Equation 3.

$$Pc = \left[\frac{N!}{A!\,(N-A)!}\right].5^{N} \qquad \text{Equation 2}$$

Pc = Occurrence probability

N = Number of experts

A = Number of experts agreeing on reasonable items

$$k^{*} = \frac{ICVI - Pc}{1 - Pc} \qquad \text{Equation 3}$$

k* = kappa approval value

ICVI = Item-level content validity index

Pc = Occurrence probability

The measurement of reasonable items was considered based on standards and criteria established by Fleiss (1971) and Cicchetti and Sparrow (1981). These standards and criteria determined whether the k* value was fair, good, or excellent as illustrated in Table 3:

Table 3. Kappa, k value*

Kappa, *k values**	**Interpretation of values**
0.40 – 0.59	Fair
0.60 – 0.74	Good
0.74 – 1.00	Excellent

Methodology

The collection of research data was carried out in three phases, namely, the administration of standard instruments, the administration of self-developed instruments, and the recording of natural data (Gay et al., 2012). Firstly, the development of assessments on knowledge and skills was designed based on standardised self-developed instruments. However, the development of assessments on knowledge and skills also depended on several factors such as the examination of the constructs, item discriminatory index, and the interpretation of the scores. Once the content validation process was made, the test instruments were piloted.

Development of test instruments

It could be recalled that the development of test instruments presented challenges as test instruments differed in terms of quality and desired goals. All test instruments were constructed based on a table of test specifications, using primary data that were collected, coded, and stored in the item bank (Mokhtar, 2016). After that, the primary data were organised based on the order of complexity. Specifically, relatively easy items were structured at the beginning of the instrument to boost confidence (a positive stimulus) before difficult items were presented to test takers.

In addition, graphics, diagrams, graphs, sketches, and tables were carefully sequenced concurrently on the same page with the instruction of assignments for clarity and response time. As soon as the test instruments were constructed, the extent to which test instrument items complied with the table of test specifications was determined. The test item instruments were reviewed by the researchers and teachers who were involved before the content validation process to ensure that the following recommendations were met (Mokhtar, 2016):

(i) High reliability

(ii) The authenticity of items (item constructions, contexts, levels of complexity, and questioning approaches)

(iii) The sequence of item difficulty (easy-to-difficult manner)

(iv) Straightforward, concise, and precise instructions

(v) Consistent structure and presentation of items (format and difficulty levels)

(vi) Compliance to tables of test specification

(vii) High readability

(viii) Explicit non-language features (organised arrangement of format, font colour, and figure colours)

(ix) Adequate time for expected responses

Assessment of knowledge

The first characteristic of the input involved knowledge. A table of test specification designed by Anderson and Krathwohl (2001) was adopted when streamlining the construction and sequence of assessment items on knowledge. Specifically, using Bloom Taxonomy Analysis as a baseline, the assessment items were carefully designed to reflect all the knowledge constructs set out in Bloom Taxonomy Analysis Model. All assessment items were drawn from a taught course called 'Control system' (CS), a course that was designed based on Vocational College Standard Curriculum (Akademik, 2017). Test takers were given one hour and 30 minutes to complete the assessment items.

The assessment test items encompassed two parts, namely, parts A and B. While part A measured areas of remembering, understanding, and application of knowledge (matching test and short-answer items), part B required test takers to analyse, evaluate, and create (data analysis), drawn based on Bloom's Taxonomy higher-order thinking questions. All items were arranged in an easy-to-difficult fashion or in the form of concrete-to-abstract reasoning questions (Krathwohl, 2002). After that, the comments on the content and item construction from the CS instructor were taken into consideration.

Table 4. Description of constructs on knowledge (Anderson & Krathwohl, 2001)

	Construct	Description
1.	Remembering	Ability to receive relevant knowledge from long-term memory
2.	Understanding	Ability to provide meaning from oral, written, and graphic instructions and messages
3.	Application	Ability to conduct or use the procedure in certain conditions
4.	Analysis	Ability to divide the material into several parts and analyse the relationship between the parts and whole structure
5.	Evaluation	Ability to make judgments based on requirements and standards
6.	Creation	Ability to bring all the elements together to form a comprehensive function and arrange the elements into a new structure

After item modifications were made, the test instruments were submitted to a panel of experts for subsequent review in terms of curriculum, contexts, and item constructions. The six constructs and construct descriptions adopted from Anderson and Krathwohl, (2001) are illustrated in Table 4.

Assessment of skills

The assessment of skills adopted Simpson's (1966) model. According to Simpson (1966), seven stages of item constructions that measured skills included perception, set, guided response, mechanisms, complex responses, adaptation, and originality. Although these seven criteria governed item constructions on learning processes and skills, the assessments on skills were administered only when the relevant skills were taught to test takers (Mokhtar, 2016). Test takers were given one hour and 30 minutes to complete the test.

The criteria as described by Simpson, (1966) guided the administration of assessments across Malaysian Vocational College through three broad constructs. Firstly, a construct on self-preparation which included an assessment at the level of perception, set, and guided response of students was considered. The second and third constructs involved work processes and student projects that satisfied the description of test items in terms of mechanisms, complex responses, adaptations, and originality. Fourthly, constructs on safety and values were considered. Specifically, constructs on values measured test takers' ability to apply knowledge, skills, and values in learning processes (Clark et al., 2010) and demonstrate test takers' ability to apply civic precepts and moral beliefs (Majid et al., 2012).

As soon as items were constructed, comments on content, item functions, and item constructions from the CS instructor were taken into consideration. After item modifications were made, the assessment on skills was submitted to a panel of experts for a subsequent review in terms of curriculum, context-specific skills, and item constructs. The four constructs and sub-constructs, including construct descriptions, are illustrated in Table 5.

Table 5. Descriptions of constructs on skills (Simpson, 1966)

	Constructs		Descriptions
1.	**Self-Preparation**	1.1 Perception	Ability to use their senses by identifying, distinguishing, and selecting before engaging in physical activities
		1.2 Set	Readiness or ability to act requiring knowledge to practise the skills
		1.3 Guided Response	Ability to mimic complex motor skills in performance. Good performance can be achieved with adequate training
2 & 3	**Work Process & Work Results**	2.1 Mechanisms	Level of confidence and explicit application of skills
		2.2 Complex Responses	Quick and accurate performance, with little to no hesitation
		2.3 Adaptation	Strong development of skills in which students can change the movement of skills to meet specific needs
		2.4 Originality	Creation of new patterns of movement to fit a particular situation or problem. The outcomes of learning emphasise on creativity
4	**Safety and Value**		Application of values and safety in tasks

Administration of content validity analysis

In this study, the content validity analysis was conducted to ensure all items were relevant and that the items only measured what the items were supposed to measure. Thus, a model by Davis (1992) on expert panel judgment governed the process of implementing instrument validation. Three processes were involved, namely, expert panel selection, expert panel consent to review test instruments, and expert panel response review results. The following descriptions highlight how these processes were carried out:

Selection of expert panels

Broad and specific criteria governed the selection of expert panels. Firstly, only expert panels who were credible in terms of their academic expertise to examine concepts, theories, and issues relevant to the content of the instrument test items were selected. Focusing on expert panels who were knowledgeable in terms of concepts, theories, and relevant issues concerning assessment test items could help to validate and produce good instruments (Davis, 1992). Secondly, specific criteria guided the expert panel selection, namely, (1) field experts whose experiences spanned more than five years of experience, (2) experts' specific experience, and (3) experts who were directly involved in the nature of the study (Akbari & Yazdanmehr, 2014). At least three experts were considered valid to review the content of assessment test items (Dimopoulos & Pantis, 2003; Lynn, 1986; Makki et al., 2003).

Therefore, five expert panels were selected to validate the content of the test instruments. Two experts who were senior lecturers teaching system controls at a large, public suburban university in the south of Peninsular Malaysia were selected to review the assessment test items. While one expert whose teaching experiences spanned over 17 years reviewed the assessment test items, another two experts with 11 years of teaching experiences at VC were also considered.

Invitation to participate in a validation process

Four steps governed the process of inviting expert panels. Firstly, all expert panels were presented with two copies of the assessment test items called 'knowledge and skill test instruments' as well as answer schemes. Secondly, letters of appointment, which detailed authorisation and consent, were distributed to expert panels. Thirdly, application forms, describing the objectives, the experts' selection criteria, and the role of experts were administered. Fourthly, an explanation of the measurement, scoring, and interpretation of the scores for all assessment test items was presented. Specifically, expert panels were made aware that scoring was based on Davis's (1992) guidelines on the use of an item evaluation scale. The five expert panels evaluated item accuracy based on a four-level Likert scale (1 = inappropriate, 2 = somewhat appropriate, 3 = appropriate and 4 = very appropriate).

Analysis on expert panels' judgment

There were four processes concerning the analysis of expert panels' judgment. Firstly, all expert panel judgments were evaluated using quantitative methods as soon as the expert panels' judgment was received. Secondly, by employing I-CVI and kappa analyses, the value of I-CVI was determined based on the expert panels' appropriate (3) and very appropriate (4) scores on the four-level Likert scale. Thirdly, by adopting the method of Polit et al. (2007), the value of I-CVI for all objects was determined. Finally, the value of k^* was calculated using the items' I-CVI values. The criteria for determining k^* values were established based on standards established by Fleiss (1971) and Cicchetti and Sparrow (1981). Thus, the value of k^* was considered 'fair' if the value fell between 0.40-0.59, 'good' if the value ranged between 0.60-0.74, and 'excellent' if the value exceeded 0.74.

Results

The expert panels' judgment revealed that items for all constructs in the assessment of knowledge were considered 'excellent', shown by the value of k^*. The results of I-CVI value analysis, k^* values, and k^* level of evaluation are illustrated in Table 6:

Table 6: Statistical analysis of modified Fleiss's kappa (assessment of knowledge)

Item	Construct	Number of Experts	Number of Expert Consent	I-CVI	*Pc*	*k**	Rating Level *k**
	Remembering Test takers were able to:						
1	Design a process system Programmable Logic Circuit (PLC)	5	5	1.00	0.031	1.00	Excellent
2	Design a process system (flow chart)	5	5	1.00	0.031	1.00	Excellent
3	Design a logic control system	5	4	0.80	0.156	0.76	Excellent
4	Develop a technical specification	5	4	0.80	0.156	0.76	Excellent
5	Produce a programming language with PLC	5	4	0.80	0.156	0.76	Excellent
6	Install a control system	5	5	1.00	0.031	1.00	Excellent
7	Maintain an uninterruptible power supply (UPS) system	5	5	1.00	0.031	1.00	Excellent

Table 6: Continued

Item	Construct	Number of Experts	Number of Expert Consent	I-CVI	*Pc*	*k**	Rating Level *k**
	Understanding Test takers were able to:						
8	Design a logic control system	5	5	1.00	0.031	1.00	Excellent
9	Develop a technical specification	5	5	1.00	0.031	1.00	Excellent
10	Produce a programming language with PLC	5	4	0.80	0.156	0.76	Excellent
11	Install a control system	5	5	1.00	0.031	1.00	Excellent
12	Execute a control system	5	5	1.00	0.031	1.00	Excellent
	Application Test takers were able to:						
13	Design a processing system	5	4	0.80	0.156	0.76	Excellent
14	Produce a programming language with PLC	5	5	1.00	0.031	1.00	Excellent
15	Execute a control system	5	4	0.80	0.156	0.76	Excellent
	Analysis, evaluation, and creation Test takers were able to:						
16	Produce a programming language with PLC	5	4	0.80	0.156	0.76	Excellent
17	Install a control system	5	5	1.00	0.031	1.00	Excellent

When it comes to assessment of skills, the expert panels' judgment stated that all items for all constructs were considered "excellent," as shown by the value of *k**. The results of I-CVI value analysis, *k** values, and *k** level of evaluation are illustrated in Table 7:

Table 7. Statistical analysis of modified Fleiss's kappa (assessment of skills)

	Construct	Number of Experts	Number of Expert Consent	I-CVI	*Pc*	*k**	Rating Level *k**
	Self-preparation (perception, set, and guided responses) Test takers were able to:						
1	Use safety protections	5	5	1.00	0.031	1.00	Excellent
2	Select information	5	5	1.00	0.031	1.00	Excellent
3	Select tools	5	5	0.80	0.156	0.76	Excellent
	Work process Test takers were able to:						
4	List details of an equipment	5	5	1.00	0.031	1.00	Excellent
5	Design a ladder diagram	5	5	1.00	0.031	1.00	Excellent
6	Build wiring	5	5	1.00	0.031	1.00	Excellent
7	Connect lights	5	5	1.00	0.031	1.00	Excellent
8	Produce programming language with PLC	5	5	1.00	0.031	1.00	Excellent
9	Test if the programme worked	5	5	1.00	0.031	1.00	Excellent
	Work outcome Test takers were able to:						
10	Demonstrate if the list of details was accurate	5	5	1.00	0.031	1.00	Excellent
11	Demonstrate if the sketch was accurate	5	4	0.80	0.156	0.76	Excellent
12	Demonstrate if the connection to PLC was accurate	5	4	0.80	0.156	0.76	Excellent
13	Demonstrate if the connection of push button (PT) was accurate	5	4	0.80	0.156	0.76	Excellent
14	Demonstrate if the connection of pilot lamp (LP) was accurate	5	4	0.80	0.156	0.76	Excellent

Table 7. Continued

	Construct	Number of Experts	Number of Expert Consent	I-CVI	*Pc*	*k**	Rating Level *k**
15	Demonstrate if the connection was organised	5	4	0.80	0.156	0.76	Excellent
16	Demonstrate if the red LP lit up for ten seconds and turned green	5	4	0.80	0.156	0.76	Excellent
17	Demonstrate if the green LP flashed ten times and went off, simultaneously demonstrating that the red LP lit up.	5	4	0.80	0.156	0.76	Excellent
18	Demonstrate if PB (start) press commenced operations	5	4	0.80	0.156	0.76	Excellent
19	Demonstrate if PB (stop) press stopped operations	5	4	0.80	0.156	0.76	Excellent
20	Clear out unnecessary materials	5	4	0.80	0.156	0.76	Excellent
21	Create a new movement	5	4	0.80	0.156	0.76	Excellent
	Safety and Values Test takers were able to:						
22	Manage equipment	5	5	1.00	0.031	1.00	Excellent
23	Organise work spaces	5	5	1.00	0.031	1.00	Excellent
24	Use personal safety and equipment	5	5	1.00	0.031	1.00	Excellent
25	Perform operations with punctuality	5	5	1.00	0.031	1.00	Excellent

Discussions

Based on the item analysis, 17 items of the assessments on knowledge were retained because the 17 items distinguished test-takers who were masters from non-masters. It was found that repeated and careful organisation and sequencing of item difficulties, including the application of Mokhtar's (2016) nine recommendations, helped construct reasonable items. In addition, the help received from experienced instructors in reviewing instruments and ensuring measurable and observable item constructions in terms of item scoring, constructs, context, and difficulty levels were considered key predictors to item analysis.

Through the suggestions received, items were modified to enhance the item discrimination index to meet the contexts and constructs of the assessment test items. With consistent organisation and execution of review by teaching instructors, item analysis was applied. As soon as modifications were made to the items following comments from teaching instructors, the finalised instrument (assessment test items) was administered, and authorisation from the selected panel of experts was sought as part of recording expert panels' judgment. Thus, the careful processes of seeking content validation from teaching instructors and expert panels yielded desirable outcomes of I-CVI and kappa analyses.

Next, 25 items of assessment of skills were also retained as an instrument that could differentiate test-takers who were masters from non-masters. Contrary to normal procedures of assessing skills, Simpson's (1966) seven recommendations (perception, set, guided response, mechanisms, complex responses, adaptation, and originality) were adopted. Comments from a VC teaching instructor concerning the adequacy of workshop equipment and context-specific skills were received and considered to ensure reasonable item constructions were built.

Following the comments received, substantial modifications were made to the assessment of skills before the assessment of skills was tested for effectiveness by some VC instructors to ensure that the assessment of skills was content and context-appropriate. When the assessment of skills was finalised as an appropriate instrument, the assessment of skills was subsequently administered to expert panels for a review on content. The findings of considerably acceptable values of I-CVI and Kappa analyses were the outcomes of rigorous processes that were prepared and carried out. In conclusion, all constructs were considered important to remove extraneous items in measuring student mastery in terms of knowledge and skills set out in the assessment of knowledge and skills.

To summarize, the findings show that the elaboration of concepts related to assessment of knowledge and assessment of skills are related to the Control System course. The VC leadership should be aware of the specific challenges and opportunities for assessments. It is to avoid educators being placed in a difficult position as professional practitioners, which will have an impact on students' development during the learning process.

Conclusion

The components of technical and vocational education through conducting content validity analysis played a crucial role in supporting the prospective highly-skilled workforce. The results of the study indicated that it was possible to identify occupational skills among students related to knowledge and skills through a variety of approaches and instruments. It should be noted, however, that content and context-specific skill assessments were identified, reviewed, and validated by expert panels to ensure only reasonable items that complied with the requirements of the syllabus and curriculum were included in assessments. By focusing on expert panels' judgment, content validation yielded an increasing trust in the context of providing valid instruments for prospective replication of research and empirical studies.

Moreover, instructors stood out in this regard because consistent, content, and context-specific assessments seemed to be related to the explicit outcome of content validity analysis. Prospective studies might better consider increasing the number of panel experts. Further research might also conduct content validity analysis involving other educational programmes because the result of the current study could only be generalised to a course on electrical technology at a Malaysian VC.

Recommendations

Based on this research, there are several recommendations to vocational lecturers or practitioners, Ministry of Education, and researchers. First, vocational lectures or practitioners are encouraged to produce final exam test or assessment according to the set standards, no matter in which field. It is recommended that a content validity test be carried out for each test question that has been completed. This is to ensure the test items are of improved quality and are equally fair to the students. Second, the Ministry of Education needs to conduct scheduled training and involve teachers, lecturers and practitioners on exposure to the production of good and quality test questions. This will make it necessary for each educator to carry out an authentic assessment and must be continually improved. Third. further research might also conduct content validity analysis involving other educational programmes because the result of the current study could only be generalised to a course on electrical technology at a Malaysian VC.

Limitations

This study only involves the development of test questions for electrical technology students who only take the Control Systems course at Malaysian Vocational College. Only vocational college lecturers and senior lecturers teaching system controls in the south of Peninsular Malaysia were involved. Researchers do not consult industry experts who may give a little more thoughtful input if taken into account.

Authorship Contribution Statement

Suhaini: Conceptualization, design, drafting manuscript, critical revision of manuscript, acquisition of data, writing, editing/ reviewing, final approval. Ahmad: Critical revision of manuscript, supervision, reviewing, final approval. Mohd Bohari: Critical revision of manuscript, editing/reviewing.

References

Agarwal, P. K. (2019). Retrieval practice & Bloom's taxonomy: Do students need fact knowledge before higher order learning? *Journal of Educational Psychology*, *111*(2), 189–209. https://doi.org/10.1037/edu0000282

Agency, M. Q. (2013). Garis panduan amalan baik: Penilaian pelajar [Good practice guidelines: Student assessment]. In *Agensi Kelayakan Malaysia* [Malaysian Qualifications Agency] (MQA). www.mqa.gov.my

Akademik, P. (2017). *Silibus: Sistem kawalan* [Syllabus: Control system]. BPTV, KPM.

Akbari, R., & Yazdanmehr, E. (2014). A critical analysis of the selection criteria of expert teachers in ELT. *Theory and Practice in Language Studies*, *4*(8), 1653–1658. https://doi.org/10.4304/tpls.4.8.1653-1658

Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy and Practice*, *25*(6), 551–575. https://doi.org/10.1080/0969594X.2018.1441807

Carmines, E. G., & Zeller, R. A. (2014). *Reliability and validity assessment*. Sage Publications.

Cicchetti, D. V., & Sparrow, S. (1981). Developing criteria for establishing interrater reliability of specific items: Application to assessment of adaptive behavior. *American Journal of Mental Deficiency*, *86*(2), 127–137.

Clark, R. W., Threeton, M. D., & Ewing, J. C. (2010). The potential of experiential learning models and practices in career and technical education & career and technical teacher education. *Journal of Career and Technical Education*, *25*(2), 46–62. https://doi.org/10.21061/JCTE.V25I2.479

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). Routledge Taylor & Francis. https://doi.org/10.4324/9780203029053

Council, N. Q. (2009). *Guide for developing assesment tools (Issue 2)*. Work-based Education Research Centre of Victoria University. https://cutt.ly/bmddLEh

Daniel, J., & Hultin, G. (2002). *Technical and vocational education and training for the twenty-first Century: UNESCO and ILO recommendations*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000126050

Davis, L. L. (1992). Instrument review: Getting the most from a panel of experts. *Applied Nursing Research*, *5*(4), 194–197. https://doi.org/10.1016/S0897-1897(05)80008-4

Dimopoulos, D. I., & Pantis, J. D. (2003). Knowledge and attitudes regarding sea turtles in elementary students on Zakynthos, Greece. *Journal of Environmental Education*, *34*(3), 30–38. https://doi.org/10.1080/00958960309603491

Education, K. D. (2013). Competency-based education: Helping all Kentucky students succeed. In *Kentucky Department of Education*. http://www.education.ky.gov

Ekonomi, U. P. (2016). *Buku rancangan Malaysia kesebelas 2016-2020* [Malaysia's eleventh plan book, 2016-2020]. Percetakan Nasional Malaysia Berhad. https://www.kkmm.gov.my/pdf/KPI/PPD KSP.pdf

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, *76*(5), 378–382. https://doi.org/10.1037/h0031619

Gay, L. R., Mills, G. E., & Airasian, P. W. (2012). *Educational research: Competencies for analysis and applications* (10th ed.). Pearson.

Ghafar, M. N. A. (1999). *Penyelidikan pendidikan* [Educational research] (Vol. 11). Penerbit Universiti Teknologi Malaysia.

Ghafar, M. N. A. (2011). *Pembinaan dan analisis ujian bilik darjah* [Developing and analysing classroom tests] (2nd ed.). Universiti Teknologi Malaysia.

Gonczi, A. (1996). *Reconceptualising competency-based education and training: With particular reference to education for occupations in Australia.* University of Technology.

Hamid, R., Baharom, S., Hamzah, N., Wan, W. H., Atiq, R., Rahmat, O. K., & Raihantaha, M. (2012). Assessment of psychomotor domain in materials technology laboratory work. *Procedia - Social and Behavioral Sciences*, *56*, 718–723. https://doi.org/10.1016/j.sbspro.2012.09.708

Jalal, W. N. A. W., Jalal, W. N. W., Sharif, A. M., & Rahman, N. A. (2017). Penggredan semula sekolah menengah vokasional kepada kolej vokasional: Kajian tahap kompetensi kaedah pengajaran dan pembelajaran pengajar [Revaluation of vocational secondary schools as vocational colleges: An investigation into the level of competence]. *SkillsMalaysia Journal*, *3*(1), 9–16.

James, M. (2010). Educational assessment: Overview. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International Encyclopedia of Education* (pp. 161–171). Elsevier.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy an overview. *Theory Into Practice*, *41*(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2

Kyobe, E., & Rugumayo, A. (2005). Competency based education and training (CBET): A case study in Uganda. In K. S (Ed.), *Proceedings of the 31st WEDC International Conference, Kampala, Uganda, 2005: Maximizing The Benefits From Water And Environmental Sanitation* (pp. 189–191). Loughborough University.

Likisa, K. D. (2018). Challenges and prospects of competency-based education: The case of adama science and technology university alumni students and hawas TVET college, Adama, Ethiopia. *The Journal of Competency-Based Education*, *3*(2), 1–7. https://doi.org/10.1002/cbe2.1163

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, *35*(6), 382–385. https://doi.org/10.1097/00006199-198611000-00017

Majid, L. A., Abdullah, W. N. W., & Zakhi, N. H. A. (2012). Penerapan nilai murni dan pembentukan jati diri kanak-kanak prasekolah melalui penggunaan multimedia [The application of values and the formation of preschool children's identity through the use of multimedia]. *Hadhari Journal / Jurnal Hadhari, 2012*(Special Edition), 51–65.

Makki, M. H., Abd-El-Khalick, F., & Boujaoude, S. (2003). Lebanese secondary school students' environmental knowledge and attitudes. *Environmental Education Research*, *9*(1), 21–33.

https://doi.org/10.1080/1350462032000034340

Malaysian Institute of Teacher Education [Institut Pendidikan Guru Malaysia]. (2019). *Buku panduan pentaksiran* [Assessment handbook]. Institut Pendidikan Guru Malaysia [Malaysian Institute of Teacher Education]. https://cutt.ly/Lmdgqkp

Malaysian Ministry of Education. (2012). *Pelan strategik interim 2011-2020* [Interim strategic plan 2011-2020]. Kememterian Pelajaran Malaysia [Malaysian Ministry of Education]. https://cutt.ly/TmdgnZV

Malaysian Ministry of Education. (2013). *Pelan pembangunan pendidikan Malaysia 2013-2025* [Malaysia's Education Development Plan 2013-2025]. Kememterian Pelajaran Malaysia [Malaysian Ministry of Education]. https://cutt.ly/wmdgAfi

Malaysian Ministry of Education. (2015). *Pelan pembangunan pendidikan Malaysia 2015-2025 (Pendidikan Tinggi)* [Malaysia's Education Development Plan 2013-2025 (Higher Education]. Kememterian Pelajaran Malaysia [Malaysian Ministry of Education]. https://cutt.ly/hmdgDK3

Mazin, K. A., Norman, H., Nordin, N., & Ibrahim, R. (2020). Student self-recording videos for TVET competency in MOOCs. *Journal of Physics: Conference Series*, *1529*, 1–8. https://doi.org/10.1088/1742-6596/1529/4/042061

McGrath, S. (2014). *Revising the 2001 revised recommendation concerning technical and vocational education*. UNESCO-UNEVOC. https://files.eric.ed.gov/fulltext/ED560511.pdf

Meyer-Adams, N., Potts, M. K., Koob, J. J., Dorsey, C. J., & Rosales, A. M. (2011). How to tackle the shift of educational assessment from learning outcomes to competencies : One program's transition. *Journal of Social Work Education*, *47*(3), 489–507. https://doi.org/10.5175/JSWE.2011.201000017

Mokhtar, M. H. (2016). *Pentaksiran pendidikan teori dan praktis ke arah transformasi sistem pendidikan negara* [Assessment of theoretical and practical education in the context of national education system transformation]. Kentalmas Jati Sdn. Bhd.

Omar, M. K., Rashid, A. M., & Puad, M. H. M. (2018). Examining job satisfaction factors toward retaining Malaysian TVET instructors in the teaching profession. *International Journal of Engineering and Technology*, *7*(2), 44–49. https://doi.org/10.14419/ijet.v7i2.10.10952

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Focus on research methods: Is the cvi an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health*, *30*, 459–467. https://doi.org/10.1002/nur.20199

Russell, M. K., & Airasian, P. W. (2011). *Classroom assessment: Concepts and applications* (7th ed.). McGraw-Hill Education.

Saha, S. K. (2021). Towards development of a system for automatic assessment of the quality of a question paper. *Smart Learning Environments*, *8*(1), 1–14. https://doi.org/10.1186/s40561-021-00148-9

Simpson, E. J. (1966). *The classification of educational objectives, psychomotor domain*. ERIC. https://files.eric.ed.gov/fulltext/ED010368.pdf

Ulum, A. R., Hidayah, N., & Yanti, Y. (2021). Development of assessment hots (higher order thinking skills) based on troubleshooting for class v sd/mi. *Journal of Madrasah Ibtidaiyah Education/ Jurnal Pendidikan Madrasah Ibtidaiyah*, *5*(1), 15–29. https://doi.org/10.32934/jmie.v5i1.222