A Content Validity Study for Vocational Teachers' Assessment Literacy Instrument (VoTAL)

Jimmy Chong, Siti Eshah Mokshein, Ramlee Mustapha

In-Text Citation: (Chong et al., 2021)
To Cite this Article: Chong, J., Mokshein, S. E., & Mustapha, R. (2021). A Content Validity Study for Vocational Teachers' Assessment Literacy Instrument (VoTAL). *International Journal of Academic Research in Business and Social Sciences*, *11*(4), 868–883.

A Content Validity Study for Vocational Teachers' Assessment Literacy Instrument (VoTAL)

Jimmy Chong[1], Siti Eshah Mokshein[1], Ramlee Mustapha[2]
[1]Faculty of Human Development, Sultan Idris Education University, Tanjong Malim, 35900, Perak, Malaysia, [2]Faculty of Technical and Vocational, Sultan Idris Education University, Tanjong Malim, 35900, Perak, Malaysia
Email: jimmychg@gmail.com

Abstract
The importance of content validity has made it a crucial step in the development of a measurement instrument. This study aimed to establish content validity for Vocational Teachers' Assessment Literacy (VoTAL) instrument using Lawshe's Content Validity Ratio (CVR) analysis. In this study, fourteen professional experts and seven lay experts were selected to assess the VoTAL instrument. A purposive sampling technique was employed to choose the expert panels based on their expertise that appropriate to the domain constructs of the instrument. The VoTAL instrument consists of three primary constructs represented by 125 items. Overall, the CVR analysis suggested that 100 items met the minimum requirements of the overall CVR value (CVR $\geq$ 0.428) and mean of judgments ($\bar{x} \geq 1.5$), whereby 25 items were excluded. In conclusion, the VoTAL instrument shows good content validity for measuring vocational teachers' assessment literacy. Hence, further study should be carried out to examine additional psychometric properties of this instrument.
Keywords: Assessment Literacy, Content Validity, Instrument Validation, Instrument Development, Experts' Judgment, Vocational Teacher, Content Validity Ratio.

Introduction
Vocational education in Malaysia has developed in numerous ways over the last decade. In 2012, all vocational schools nationwide were upgraded to Vocational College to meet the 11th Malaysia Plan agenda, which is expected to create about 1.5 million job opportunities, with 60 percent of these jobs coming from TVET-related sectors (Economic Planning Unit, 2015). The development in vocational education not only brought changes in the teaching and learning system but also improvements in the assessment of vocational classrooms. Vocational College employs two types of assessments. The first type of assessment is centralized and contributes to 30 percent of overall assessments. The Malaysian Examination Syndicate administers this centralized assessment. The second type of assessment is competency-based and contributes to 70 percent of the overall assessments. Teachers in classrooms or vocational workshops perform this assessment. Therefore, due to the importance of teachers' role in the assessment of student's achievement, the demand for

accountability and quality of assessment has become an issue in vocational education (Mohamed et al., 2016; Rahman et al., 2014; Retnawati et al., 2016).

Classroom assessment has been described as the cornerstone of the current education system (Brookhart, 2011; Kelly et al., 2020; Popham, 2014). Previous researches had emphasized the significance of assessment in strengthening the education system (Birenbaum et al., 2015; Brookhart, 2011; DeLuca, 2012; Popham, 2014). Additionally, constant inclusion of student assessments in the teaching and learning process has demonstrated improvements in students' achievement, motivation, positive self-esteem and improve teachers' instruction (Bennett, 2011; DeLuca & Klinger, 2010; Harlen, 2012; Maclellan, 2004; Willis et al., 2013).

In view of these growing benefits, the attention on supporting teachers' assessment literacy has increased (Deluca et al., 2016a). Thus, the Vocational Teachers' Assessment Literacy (VoTAL) instrument had been developed to assess vocational teachers' self-perceived assessment literacy. In order to enhance the instrument's validity, the newly developed VoTAL instrument is subjected to a content validity test. The content validity of the VoTAL instrument was analyzed by professional and lay expert panels. The data collected from the expert panels were analyzed by using Lawshe's Content Validity Ratio (CVR) method.

Literature Review
Assessment literacy
Teachers' assessment literacy is teachers' capability to express classroom and social culture perspectives to implement, create, perform, incorporate, and apply effective assessment strategies to supports students' learning against education standards (Alonzo, 2020; Willis et al., 2013). Given the persistent demands to improved and enhanced teachers' assessment literacy (Popham, 2016; Shepard, 2020), previous studies had consistently shown that teachers experience significant challenges in implementing contemporary assessment methods that reflect the current assessment landscape requirements (DeLuca et al., 2018). Consequently, teachers maintain a differing perspective of educational assessment that leads to differences in classroom assessment practices. Studies also showed little credible information on teachers' assessment literacy concerning the current accountability expectations caused by inadequate contemporary assessment literacy measurement scales that predicted the present educational assessment standards (DeLuca et al., 2016a). This finding is not unexpected since most of the assessment literacy instruments are predicted on the early 1990s standard, which is the Standards for Teacher Competence in Educational Assessment of Students (STCEAS) (Deluca et al., 2016a; Gotch & French, 2014). Although the STCEAS 1990 standard (American Federation of Teachers [AFT], National Council on Measurement in Education [NCME], & National Educational Association [NEA], 1990) had contributed significantly to the research on assessment literacy, Brookhart (2011) states that the STCEAS 1990 assessment standard is outdated in two aspects; (1) it does not take into consideration the current concept of formative assessment; and (2) it also does not take into consideration the social issues, as well as the different aspects teachers, encounter in contemporary assessment landscape.

To reflect on these issues, recently published Classroom Assessment Standards by Joint Committee on Standards for Educational Evaluation – JCSEE 2015 (Klinger et al., 2015) has addressed this criticism by stating a set of guidelines and principles in the assessment of students' learning, which in line with the current context of educational assessment landscape. The evolution of classroom assessment requirements, together with the

development of the JCSEE 2015 standard, indicates that previous measurement scales to assess teachers' assessment literacy may not provide strong validity concerning the present contemporary classroom assessment context (Deluca et al., 2016a; Gotch & French, 2014). Thus, in an attempt to expand research in assessment literacy beyond the STCEAS 1990 standards, this study had developed the Vocational Teachers' Assessment Literacy (VoTAL) instrument. The VoTAL instrument was developed based on the newly published JCSEE 2015 standard. This standard provides contemporary alternatives to the STCEAS 1990 standards and reflects accurately the assessment expectations of teachers within the accountability driven education systems (DeLuca et al. 2016b).

Content Validity

Content validity is an essential aspect that needs to be addressed in developing a new measurement instrument. DeVellis (2003) stressed that content validity is the first kind of validity that will be assessed in developing a new instrument. Content validity is the extent to how each item in an instrument is suitable and represents the definition of a domain construct (DeVellis, 2003; Furr, 2011; Rubio et al., 2003). Content validity is an assessment for each item to make them appropriate for the instrument development purpose. An instrument with strong content validity should include only relevant and essential items that appropriately address the instrument's construct (Nunnally & Bernstein, 1994). Content validity is critical to ensure the measurement instrument is measuring what it should be measured. Thus, this study highlighted one of the crucial phases in the development of a new instrument by demonstrated the method to determine the content validity of the VoTAL instrument. The Content Validity Ratio (CVR) method is employed to assess the content validity in this study.

CVR is a broadly employed method to measure the content validity of an instrument through empirical assessment. The CVR method, which was developed by Charles Lawshe in 1975, helps researchers to determine whether to keep or remove items on measurement instruments through CVR calculations. In short, it aims to filter items empirically with quantitative procedures to ensure that each item truly represents the content of the domain construct. CVR employs a group of expert panels to examine the degree to which each item reflects the domain construct of an instrument. The group of experts will assess each item on the three-point scale, (1) essential, (2) useful but not essential, and (3) not essential. The CVR method is selected in this study as it is practical in terms of time and cost, straight forward, simple, and easy to apply. Apart from that, CVR also provides a table for determining the critical cut-off value and emphasizing the statistical significance of agreement at the item level.

Methods
The VoTAL Instrument

The VoTAL instrument was developed with the aim to assess vocational teachers' self-perceived assessment literacy. This instrument was developed based on the JCSEE 2015 standard and consisted of 125 items to represent three primary constructs and 14 sub-constructs. The assessment literacy constructs covered in this instrument are (1) Assessment Foundation, (2) Use of Assessment, and (3) Assessment Quality. Table 1 shows the items' distribution for the VoTAL instrument.

Table 1

Items' distribution for the VoTAL instrument

Constructs	Sub-Constructs	Item Number	Total Item
Assessment Foundation	Assessment Purpose	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	10
	Learning Expectations	11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21	11
	Assessment Design	22, 23, 24, 25, 26, 27, 28, 29, 30, 31	10
	Student Engagement in Assessment	32, 33, 34, 35, 36, 37, 38	7
	Assessment Preparation	39, 40, 41, 42, 43, 44, 45, 46	8
	Informed Students and Parents/ Guardians	47, 48, 49, 50, 51, 52	6
Use of Assessment	Analysis of Student Performance	53, 54, 55, 56, 57, 58, 59	7
	Effective Feedback	60, 61, 62, 63, 64, 65, 66, 67, 68	9
	Instructional Follow-Up	69, 70, 71, 72, 73, 74, 75, 76	8
	Grades and Summary Comments	77, 78, 79, 80, 81, 82, 83, 84, 85	9
	Reporting	86, 87, 88, 89, 90, 91, 92	7
Assessment Quality	Unbiased and Fair Assessment	93, 94, 95, 96, 97, 98, 99,100, 101, 102, 103, 104, 105, 106	14
	Reliability and Validity	107, 108, 109, 110, 111, 112, 113, 114, 115	9
	Reflection	116, 117, 118, 119, 120, 121, 122, 123, 124, 125	10
		Total Item	125

Identification of Validation Panels

Experts are individuals who possess the knowledge and expertise in a specific field (Nur Farhana et al., 2018). Selecting the right validation panels may influence the reliability of the validation process in determining if the measurement instrument is appropriately developed and suited to psychometric assessment (Grant & Davis, 1997). The validation panels in this study consist of professional and lay experts. The professionals in this study were selected from the practitioners with research or fieldwork experience. To allow the representation of the study population for the developed instrument, the lay experts were chosen from the potential subject of the study (Rubio et al., 2003; Zamanzadeh et al., 2015). The professional experts were identified based on their background in the research area with specific expertise, possessing the related working experience, and having up-to-date knowledge, while the lay experts were selected for their work in the appropriate field (Nur Farhana et al., 2018; Powell, 2003; Rubio et al., 2003).

Determining the number of expert panels has always been diverse in the literature. As suggested by Zamanzadeh (2015), a minimum of five experts is required to achieve adequate control on the possible agreement. Meanwhile, Lynn (1986) recommended a minimum of three experts, while Rubio et al. (2003) suggest for each category of experts to have between

three to ten panels. Other scholars such as Gable and Wolf (1993) and Tilden et al. (1990) had proposed two to twenty panels for each category of experts. The Lawshe method of content validation only requires a minimum of four experts. However, more information about the measure can be obtained by using a larger number of experts (Allahyari et al., 2009; Gable & Wolf, 1993; Rubio et al., 2003). Even so, the final decision on the number of panel experts is based on the experts' level of competence, knowledge, and depth of experience (Grant & Davis, 1997).

Expert Panels
In this study, a purposive sampling technique was employed to select 14 professional experts and seven lay experts. The experts were chosen according to their expertise appropriate to the domain constructs of the instrument. The expert panels were contacted in person to obtain their agreement to take part in the study. Before their approval, the expert panels will be informed about the study purpose, the reason for them being selected, and the experts' role in the study. Appointment letters and the validation materials were sent by email. The expert panels were provided with detailed instructions for validating the instrument. The expert panels were also requested to identify areas of deficiency and provide recommendations to improve the sentence structure and clarity of the items. All the expert panels were given two weeks to assess and validate 125 items in the VoTAL instrument. This study employs an online method for the assessment and validation process as the selected experts were from a different and large geographical location.

Lawshe's CVR Model Modification
CVR employs a group of expert panels to assess the appropriateness of an instrument's items reflect the domain construct on the three-point scale, (1) essential, (2) useful but not essential, and (3) not essential. However, there have been some criticisms of Lawshe's CVR model in assessing the agreement and response of the panels (Ahmad et al., 2019; Allahyari et al., 2009; Chalavi et al., 2015). Thus, to prevent different misunderstandings related to Lawshe's codes and to provide more significant differentiation in panels' ratings, Lawshe's three-point rating scales were expanded to a five-point scale (Ahmad et al., 2019; Allahyari et al., 2009; Chalavi et al., 2015). As to provide the panels with three different options as suggested by Leedy and Ormrod (2016) guideline, the five-point Likert scale for the judgments is composed of two positive, one neutral, and two negative scales. In comparison to Lawshe's scale, the proposed five-point Likert scale is better as it provides a wider selection range and clear sentences (Ahmad et al., 2019; Allahyari et al., 2009). Therefore, the expert panels for this study were instructed to provide their judgments regarding the suitability of each item to the domain construct based on the defined five-points scale, which includes, 1= totally not suitable; 2= not suitable; 3 = less suitable; 4 = suitable; 5 = very suitable.

Quantifying of Consensus Among Panellists
The consensus of judgments among expert panels on the necessity to include an item in the measure can be quantified by determining the content validity ratio (CVR). The judgments of expert panels who made the suitable (4) and very suitable (5) selection were computed using the content validity ratio formula:

$$CVR = \frac{n_e - (N/2)}{N/2}$$

The n_e from the above formula denotes the number of panels which made the suitable (4) and very suitable (5) selection, while N refers to the total panels. The outcome of this formula can be described that the CVR value is closer to value 1 when all panels agreed that the item is suitable (4) and very suitable (5). The CVR values ranged from 0 to 1 when over half of the panels made the suitable (4) and very suitable (5) selection. The CVR value is negative (CVR <0) when less than half of the panels made the suitable (4) and very suitable (5) selection. The CVR value acceptance criteria on items are based on the revised version of the reference table (Wilson et al., 2012), which was initially developed by Lawshe (1975). The revised version of the reference table is shown in Table 2.

Table 2
Acceptance CVR Values Based on Lawshe (1975) and Reviewed by Wilson et al. (2012)

N	Level of Significance for One-Tailed Test					
	.1	.05	.025	.01	.005	.001
	Level of Significance for Two-Tailed Test					
	.2	.1	.05	.02	.01	.002
5	.573	.736	.877	.99	.99	.99
6	.523	.672	.800	.950	.99	.99
7	.485	.622	.741	.879	.974	.99
8	.453	.582	.693	.822	.911	.99
9	.427	.548	.653	.775	.859	.99
10	.405	.520	.620	.736	.815	.977
11	.387	.496	.591	.701	.777	.932
12	.370	.475	.566	.671	.744	.892
13	.356	.456	.544	.645	.714	.857
14	.343	.440	.524	.622	.688	.826
15	.331	.425	.506	.601	.665	.798
16	.321	.411	.490	.582	.644	.773
17	.311	.399	.475	.564	.625	.750
18	.302	.388	.462	.548	.607	.729
19	.294	.377	.450	.534	.591	.709
20	.287	.368	.438	.520	.576	.691
21	.280	.359	.428	.508	.562	.675
22	.273	.351	.418	.496	.549	.659
23	.267	.343	.409	.485	.537	.645
24	.262	.336	.400	.475	.526	.631
25	.256	.329	.392	.465	.515	.618
26	.251	.323	.384	.456	.505	.606
27	.247	.317	.377	.448	.496	.595
28	.242	.311	.370	.440	.487	.584
29	.238	.305	.364	.432	.478	.574
30	.234	.300	.358	.425	.470	.564
31	.230	.295	.352	.418	.463	.555
32	.227	.291	.346	.411	.455	.546
33	.223	.286	.341	.405	.448	.538
34	.220	.282	.336	.399	.442	.530
35	.217	.278	.331	.393	.435	.522
40	.203	.260	.310	.368	.407	.489

Based on Table 2, with the total number of 21 expert panels, the CVR critical value for each item must be equal to or greater than 0.428 at α = .05 level of significance for a two-

tailed test. Any item with the CVR critical value less than 0.428 will be excluded from the instrument.

Calculation of The Respective Judgments' Means ($\bar{x}$)

To calculate the mean of judgments ($\bar{x}$) for every item, the values reflected in the instrument will be converted according to the following conversion rules (Allahyari et al., 2009; Chalavi et al., 2015):

- Very Suitable or Suitable - was replaced with 2
- Less Suitable - was replaced with 1
- Totally Not Suitable or Not Suitable - was replaced with 0

The total value of Very Suitable (2), Suitable (1), and Less Suitable (0) for every item is summed up and divided with the total number of panels. The items which did not meet the minimum values were refined or considered to be excluded from the final instrument.

Determination of Items' Acceptance and Rejection Criteria

The following requirements were used for the determination of acceptance and rejection of items in this instrument.

(1) An item is unconditionally accepted if CVR ≥ 0.428. This value is determined by the number of expert panels (N = 21, Table 2).

(2) An item is accepted if CVR value is 0 < CVR < 0.428 and $\bar{x}$ ≥1.5. The CVR value between zero and 0.428 indicates that over half of the expert panels rate the items as "Suitable or Very Suitable." The $\bar{x}$ ≥1.5 shows that the mean of judgments is close to the value of "Suitable or Very Suitable" and indicates that the mean is equal to or greater than 75 percent of the maximum mean value (2), which is higher than the minimum acceptable value of 60 percent (Chalavi et al., 2015).

(3) An item is either refined or excluded from the instrument if CVR ≤ 0 and $\bar{x}$ < 1.5. These values show that the item was not judged to be "Suitable or Very Suitable" (essential item in Lawshe's scale) by at least half of the panel and possessing a mean of judgments that is closer to "Totally Not Suitable or Not Suitable" (unessential item in Lawshe's scale).

Results

A total of 21 expert panels, which include 14 professional experts and seven lay experts, were involved in validating the VoTAL instrument. The response rate received from all the professional and lay expert panels was 100 percent. All the expert panels (professional and lay experts) completed their assessment within the given period. The 14 professional expert panels were academicians who worked as lecturers or researchers in the education sector. In contrast, the seven lay expert panels were the research subject who directly involved with the assessment process in the vocational college. The list of professional and lay expert panels' expertise and years of experience are shown in table 3 and table 4, respectively. Based on table 3, the 14 professional experts' years of experience were ranged from five to 38 years. While based on table 4, the seven lay expert panels' years of experience were ranged from 11 to 29 years.

Table 3
List of professional expert panels

Professional Expert Code	Expertise	Experience (Years)
P1	Educational measurement, Technical and Vocational education	18
P2	Educational measurement	38
P3	Technical and Vocational education	5
P4	Educational measurement	15
P5	Technical and Vocational education	22
P6	Technical and Vocational education	10
P7	Educational measurement	14
P8	School assessment	7
P9	Psychometric, Technical and Vocational education	10
P10	Educational measurement	12
P11	Technical and Vocational education, Teacher's Competency	11
P12	Technical and Vocational education	12
P13	Education assessment, Technical and Vocational education	14
P14	Technical and Vocational education, Engineering education	12

Table 4
List of lay expert panels

Lay Experts Code	Expertise	Experience (Years)
L1	Technical and Vocational education, Excellent Vocational Teacher	15
L2	Technical and Vocational education	25
L3	Technical and Vocational education, Problem based learning	11
L4	Technical and Vocational education curriculum	17
L5	Vocational assessment	29
L6	Technical and Vocational education	22
L7	Technical and Vocational education curriculum	11

Table 5 shows the results of the CVR analysis for each item based on 14 professional experts, seven lay experts, and a total of 21 experts. However, item rejection and acceptance criteria were based solely on the overall CVR value of 21 expert panels and the mean of judgments for each item. The CVR value for the professional and lay experts is shown only for comparison purposes. Based on the CVR acceptance values shown in Table 2, the CVR requirements for the total of 21 expert panels are CVR ≥ 0.428. Apart from that, the mean of judgments value ($\bar{x}$) must be equal to or greater than 1.5. Thus, the overall analysis of 125 items showed that 25 items did not comply with the minimum requirement of the overall CVR value (CVR ≥ 0.428) and mean of judgments ($\bar{x}$ ≥1.5). These items are item number 10, 19, 20, 21, 24, 27, 32, 44, 48, 67, 68, 76, 81, 88, 95, 96, 97, 98, 103, 104, 115, 121, 123, 124 and 125. All of these items were excluded from the instrument.

Table 5

CVR, means of judgments, and acceptance/rejection results

Item Number	Total (N=21) CVR ≤0.428	Mean of Judgments ($\bar{x}$) <1.50	Item Status	Item Number	Total (N=21) CVR ≤0.428	Mean of Judgments ($\bar{x}$) <1.50	Item Status
1	0.714	1.81	Accepted	27	**0.238**	**1.43**	**Rejected**
2	0.714	1.86	Accepted	28	0.905	1.90	Accepted
3	0.905	1.90	Accepted	29	1.000	2.00	Accepted
4	0.714	1.81	Accepted	30	1.000	2.00	Accepted
5	0.810	1.90	Accepted	31	0.905	1.90	Accepted
6	0.810	1.90	Accepted	32	**0.238**	**1.48**	**Rejected**
7	0.905	1.95	Accepted	33	0.524	1.71	Accepted
8	0.524	1.71	Accepted	34	0.905	1.95	Accepted
9	0.810	1.90	Accepted	35	0.810	1.86	Accepted
10	**0.048**	**1.10**	**Rejected**	36	0.714	1.86	Accepted
11	0.714	1.71	Accepted	37	0.905	1.95	Accepted
12	0.524	1.71	Accepted	38	0.714	1.86	Accepted
13	0.714	1.76	Accepted	39	0.810	1.81	Accepted
14	0.524	1.62	Accepted	40	0.810	1.86	Accepted
15	0.810	1.86	Accepted	41	1.000	2.00	Accepted
16	0.714	1.81	Accepted	42	0.810	1.86	Accepted
17	0.524	1.71	Accepted	43	0.905	1.90	Accepted
18	0.714	1.81	Accepted	44	**0.143**	**1.43**	**Rejected**
19	**-0.048**	**1.05**	**Rejected**	45	0.905	1.90	Accepted
20	**0.143**	**1.24**	**Rejected**	46	0.810	1.90	Accepted
21	**0.333**	**1.48**	**Rejected**	47	0.810	1.90	Accepted
22	0.714	1.81	Accepted	48	0.714	1.81	Accepted
23	0.619	1.76	Accepted	49	0.429	1.71	Accepted
24	**-0.048**	**1.05**	**Rejected**	50	0.714	1.81	Accepted
25	0.905	1.90	Accepted	51	0.810	1.90	Accepted
26	0.619	1.71	Accepted	52	0.714	1.81	Accepted
53	1.000	2.00	Accepted	90	0.810	1.90	Accepted
54	0.524	1.76	Accepted	91	0.619	1.76	Accepted
55	0.524	1.71	Accepted	92	0.714	1.76	Accepted
56	0.429	1.62	Accepted	93	0.905	1.90	Accepted
57	0.905	1.95	Accepted	94	0.714	1.81	Accepted
58	0.714	1.86	Accepted	95	**-0.048**	**1.00**	**Rejected**
59	0.905	1.95	Accepted	96	**0.143**	**1.43**	**Rejected**
60	0.905	1.90	Accepted	97	**-0.048**	**1.24**	**Rejected**
61	0.905	1.95	Accepted	98	**0.238**	**1.33**	**Rejected**
62	0.905	1.90	Accepted	99	0.524	1.76	Accepted
63	0.810	1.86	Accepted	100	0.524	1.62	Accepted
64	**0.238**	**1.38**	**Rejected**	101	0.810	1.86	Accepted
65	1.000	2.00	Accepted	102	0.619	1.67	Accepted
66	0.905	1.95	Accepted	103	**0.333**	**1.43**	**Rejected**

67	**0.333**	**1.43**	**Rejected**	104	**0.143**	**1.38**	**Rejected**
68	**0.238**	**1.29**	**Rejected**	105	0.905	1.90	Accepted
69	0.810	1.90	Accepted	106	0.619	1.67	Accepted
70	1.000	2.00	Accepted	107	0.905	1.90	Accepted
71	0.905	1.90	Accepted	108	0.714	1.81	Accepted
72	1.000	2.00	Accepted	109	0.810	1.81	Accepted
73	0.524	1.57	Accepted	110	0.714	1.71	Accepted
74	1.000	2.00	Accepted	111	0.905	1.95	Accepted
75	1.000	2.00	Accepted	112	0.905	1.95	Accepted
76	**-0.048**	**1.10**	**Rejected**	113	0.714	1.81	Accepted
77	0.810	1.90	Accepted	114	0.619	1.71	Accepted
78	0.714	1.81	Accepted	115	**0.333**	**1.38**	**Rejected**
79	0.905	1.95	Accepted	116	0.905	1.90	Accepted
80	0.905	1.95	Accepted	117	0.810	1.86	Accepted
81	**0.143**	**1.48**	**Rejected**	118	1.000	2.00	Accepted
82	0.810	1.90	Accepted	119	0.905	1.90	Accepted
83	1.000	2.00	Accepted	120	0.714	1.86	Accepted
84	0.905	1.95	Accepted	121	**0.048**	**1.14**	**Rejected**
85	0.524	1.62	Accepted	122	1.000	2.00	Accepted
86	0.905	1.95	Accepted	123	**0.143**	**1.48**	**Rejected**
87	0.905	1.95	Accepted	124	**0.143**	**1.38**	**Rejected**
88	**0.048**	**1.19**	**Rejected**	125	**0.238**	**1.43**	**Rejected**
89	0.619	1.76	Accepted				

The 25 items that were not fulfilled the requirements were from the Assessment Foundation (9 items), Use of Assessment (5 items), and Assessment Quality (11 items) constructs. Table 6 shows the rejected items and their respective constructs.

Table 6
The constructs and rejected items

Constructs	Sub-Constructs	Item Number	CVR ≤ 0.428	$\bar{x}<1.50$	Total Rejected Item
Assessment Foundation	Assessment Purpose	10	0.048	1.10	9
	Learning Expectations	19	-0.048	1.05	
		20	0.143	1.24	
		21	0.333	1.48	
	Assessment Design	24	-0.048	1.05	
		27	0.238	1.43	
	Student Engagement in Assessment	32	0.238	1.48	
	Assessment Preparation	44	0.143	1.43	
	Informed Students and Parents/ Guardians	48	0.238	1.38	
Use of Assessment	Effective Feedback	67	0.333	1.43	5
		68	0.238	1.29	
	Instructional Follow-Up	76	-0.048	1.10	
	Grades and Summary Comments	81	0.143	1.48	
	Reporting	88	0.048	1.19	
Assessment Quality	Unbiased and Fair Assessment	95	-0.048	1.00	11
		96	0.143	1.43	
		97	-0.048	1.24	
		98	0.238	1.33	
		103	0.333	1.43	
		104	0.143	1.38	
	Reliability and Validity	115	0.333	1.38	
	Reflection	121	0.048	1.14	
		123	0.143	1.48	
		124	0.143	1.38	
		125	0.238	1.43	
				Total rejected item	25

Discussions

VoTAL instrument is designed to assess vocational teachers' self-perceived assessment literacy. This study had established the VoTAL instrument's content validity by employing Lawshe's (1975) CVR model. To improve the instrument's validity, a joint panel consisted of both professional experts and lay experts were invited. In addition, due to the confusion related to Lawshe's codes from the previous study, the Likert scale was used to improve the response process. As suggested by Rubio et al. (2003), all expert panels in this study were analyzed together and did not distinguish between professional and lay experts. Although the use of expert panels provides valuable information for the researcher to revise the instrument, there are also some limitations to this method. Expert input feedback is

subjective; thus, analyses are subject to bias that can occur amongst experts (Zamanzadeh et al., 2015).

The final results of CVR calculations show that 25 items below the minimum requirement criteria of CVR ≥ 0.428 and $\bar{x}$ ≥1.5. Only 100 items were accepted for the final instrument. The accepted items show CVR values are between 0.428 to 1.00, and the mean of judgments ranged from 1.57 to 2.00. Twelve items showed a 100 percent consensus from the expert panels with the CVR value of 1.00 and the mean of judgments of 2.00. The accepted items were then arranged according to the format specified. Besides, some of the accepted items were also revised to enhance the wording and sentence structure. Thus, the refinement and improvement processes were made for the final 100 items. Comments and recommendations from the expert panels on every item were analyzed critically and taken into consideration to ensure the quality of each item. Measurement of assessment literacy requires a valid and reliable instrument. Thus, the findings of this study showed that experts in the field of assessment literacy believed that the selected items had the potential to be used for the said purpose. The findings from this study also indicated that the VoTAL instrument is a promising instrument that could be used to assess self-perceived assessment literacy among vocational teachers.

Conclusions

This study illustrates how to perform a content validity analysis by using Lawshe's CVR method, which is a critical step in instrument development. Expert panels were used to assess and judge the instrument items. This new VoTAL instrument has demonstrated adequate and acceptable content validity. Based on the calculated CVR value of 125 items from the 21 expert panels, only 25 items were below the requirements' criteria. In conclusion, after the content validation process, a total of 100 items from 125 items remain. The CVR is a prominent measurement model that quantifies agreement among experts through statistical analysis. The decision to include or exclude items was made clearly and appropriately. Apart from that, the psychometric properties of measuring instruments must also be rigorously tested. Therefore, for the better applicability of the measurement instrument, future research should warrant that each instrument is subjected to appropriate validity tests. Thus, the revised version of the VoTAL instrument will be subjected to a pilot study to investigate other additional validity and psychometric properties of the instrument.

Overall, the content validity study on the VoTAL instrument has contributed to the existing literature of instrument validity research by demonstrating the used of Lawshe's CVR approach method to assess the content validity of an instrument. Apart from that, this study has also contributed to the existing assessment literacy research through the development and validation of the VoTAL instrument, which was develop based on the current JCSEE 2015 classroom assessment standard. The VoTAL instrument was designed to assess vocational teachers' self-perceived assessment literacy. Several instruments had been established over the years to assess teachers' assessment literacy, such as Classroom Assessment Literacy Inventory (Mertler, 2003), Assessment Literacy Inventory (Campbell et al., 2002), Teacher Assessment Literacy Questionnaire (Plake, 1993) and Assessment in Vocational Classroom Questionnaire (Kershaw, 1993). Although progress has been made in developing valid and reliable assessment literacy measures, this VoTAL instrument is different from others as it was developed based on the new JCSEE 2015 assessment standard rather than the STCEAS 1990 standards. This study extends the work of DeLuca et al. (2016a) by promoting the use of the JCSEE 2015 standard in dealing with current contemporary assessment demand.

References

Ahmad, N. A., Drus, S. M., Kasim, H., & Othman, M. M. (2019). Assessing Content Validity of Enterprise Architecture Adoption Questionnaire (EAAQ) Among Content Experts. *2019 IEEE 9th Symposium on Computer Applications & Industrial Electronics (ISCAIE),* 160–165. https://doi.org/10.1109/ISCAIE.2019.8743918

Allahyari, T., Rangi, N. H., Khosravi, Y., & Zayeri, F. (2009). Development and evaluation of a new questionnaire for rating of cognitive failures at work. *International Journal of Occupational Hygiene*, *3*(1) 6–11.

Alonzo, A. C. (2020). Defining Trustworthiness for Teachers' Multiple Uses of Classroom Assessment Results. In Susan M. Brookhart & J. H. Mcmillan (Eds.), *Classroom Assessment and Educational Measurement* (pp. 120–145). Taylor & Francis.

American Federation of Teachers [AFT], National Council on Measurement in Education [NCME], & National Educational Association [NEA]. (1990). *Standards for Teacher Competence in Educational Assessment of Students*. National Council on Measurement in Education.

Bennett, R. E. (2011). Formative Assessment: A Critical Review. *Assessment in Education: Principles, Policy & Practice*, *18*(1), 5–25. https://doi.org/10.1080/0969594X.2010.513678

Birenbaum, M., DeLuca, C., Earl, L., Heritage, M., Klenowski, V., Looney, A., Smith, K., Timperley, H., Volante, L., & Wyatt-Smith, C. (2015). International trends in the implementation of assessment for learning: Implications for policy and practice. *Policy Futures in Education, 13*(1), 117–140. https://doi.org/10.1177/1478210314566733

Brookhart, S. M. (2011). Educational Assessment Knowledge and Skills for Teachers. *Educational Measurement: Issues and Practice*, *30*(1), 3–12. https://doi.org/10.1111/j.1745-3992.2010.00195.x

Campbell, C., Murphy, J. A., & Holt, J. K. (2002). Psychometric analysis of an assessment literacy instrument: Applicability to preservice teachers. *Annual Meeting of the Mid-Western Educational Research Association, Columbus, OH.*

Chalavi, L. H., Alamdari, S., & Asadi, K. (2015). Improving Indices in Evaluation of Medical Research Centers in Iran. *International Journal of Advanced Research in Biological Sciences*, *2*(2), 41–52.

DeLuca, C. (2012). Preparing Teachers for the Age of Accountability: Toward a Framework for Assessment Education. *Action in Teacher Education*, *34*(5–6), 576–591. https://doi.org/10.1080/01626620.2012.730347

DeLuca, C., & Klinger, D. A. (2010). Assessment Literacy Development: Identifying Gaps in Teacher Candidates' Learning. *Assessment in Education: Principles, Policy & Practice*, *17*(4), 419–438. https://doi.org/10.1080/0969594X.2010.516643

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016a). Approaches to Classroom Assessment Inventory: A New Instrument to Support Teacher Assessment Literacy. *Educational Assessment*, *21*(4), 248–266. https://doi.org/10.1080/10627197.2016.1236677

DeLuca, C., Valiquette, A., Coombs, A., LaPointe-McEwan, D., & Luhanga, U. (2016b). Teachers' approaches to classroom assessment: a large-scale survey. *Assessment in Education: Principles, Policy & Practice*, *25*(4), 355–375. https://doi.org/10.1080/0969594X.2016.1244514

DeVellis, R. F. (2003). *Scale Development: Theory and Applications*. SAGE Publications.

Economic Planning Unit. (2015). 11th Malaysia Plan (2016-2020). In *Prime Minister Department*. http://www.epu.gov.my

Furr, M. (2011). *Scale Construction and Psychometrics for Social and Personality Psychology*. SAGE Publications.

Gable, R. K., & Wolf, M. B. (1993). *Instrument Development in the Affective Domain* (2nd ed.). Springer Netherlands.

Gotch, C. M., & French, B. F. (2014). A Systematic Review of Assessment Literacy Measures. *Educational Measurement: Issues and Practice*, *33*(2), 14–18. https://doi.org/10.1111/emip.12030

Grant, J. S., & Davis, L. L. (1997). Selection and use of content experts for instrument development. *Research in Nursing & Health*, *20*(3), 269–274. https://doi.org/10.1163/15718182-02502011

Harlen, W. (2012). On the Relationship between Assessment for Formative and Summative Purposes. In J. Gardner (Ed.), *Assessment and Learning* (2nd ed.). SAGE Publications. https://doi.org/10.4135/9781446250808

Kelly, M. P., Feistman, R., Dodge, E., St. Rose, A., & Littenberg-Tobias, J. (2020). Exploring the dimensionality of self-perceived performance assessment literacy (PAL). *Educational Assessment, Evaluation and Accountability*, *32*(4), 499–517. https://doi.org/10.1007/s11092-020-09343-7

Kershaw, I. (1993). *Ohio vocational education teachers' perceived use of student assessment information in educational decision-making*. The Ohio State University.

Klinger, D., J. McDivitt, P., B. Howard, B., A. Munoz, M., Roger, T. W., & Wylie, E. (2015). *Classroom Assessment Standards for PreK-12 Teachers*. Joint Committee on Standards for Educational Evaluation.

Lawshe, C. H. (1975). A Quantitative Approach to Content Validity. *Personnel Psychology*, *28*(4), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Leedy, P. D., & Ormrod, J. E. (2016). *Practical Research: Planning and Design*. Pearson Education.

Lynn, M. R. (1986). Determination and Quantification of Content Validity. *Nursing Research*, *35*(6), 382–386.

Maclellan, E. (2004). Initial knowledge states about assessment: novice teachers' conceptualizations. *Teaching and Teacher Education*, *20*(5), 523–535. https://doi.org/10.1016/j.tate.2004.04.008

Mertler, C. A. (2003). *Preservice versus inservices teachers' assessment literacy: Does classroom experience make a difference?* Paper presented at the Annual Meeting of the Mid-Western Educational Research Association, Columbus, OH.

Mohamed, S., Kamis, A., & Ali, N. (2016). Gauging the assessment literacy of Malaysia's home economics teachers: An empirical study. *Geografia-Malaysian Journal of Society and Space*, *3*(3), 130–138.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. McGraw-Hill.

Plake, B. S. (1993). Teacher Assessment Literacy: Teachers' Competencies in the Educational Assessment of Students. *Mid-Western Educational Researcher*, *6*(1), 21-27.

Popham, W. J. (2014). *Classroom Assessment: What Teachers Need to Know*. Pearson Education.

Popham, W. J. (2016). *Classroom Assessment: What Teachers Need to Know*. Pearson Education.

Powell, C. (2003). The Delphi technique: myths and realities. *Journal of Advanced Nursing*,

41(4), 376–382. https://doi.org/10.1046/j.1365-2648.2003.02537.x

Rahman, A. A., Hanafi, N. M., Mukhtar, M. I., & Ahmad, J. (2014). Assessment Practices for Competency based Education and Training in Vocational College, Malaysia. *Procedia - Social and Behavioral Sciences*, *112*, 1070–1076. https://doi.org/10.1016/j.sbspro.2014.01.1271

Retnawati, H., Hadi, S., & Nugraha, A. C. (2016). Vocational High School Teachers' Difficulties in Implementing the Assessment in Curriculum 2013 in Yogyakarta Province of Indonesia. *International Journal of Instruction, 9*(1), 33–48. https://doi.org/10.12973/iji.2016.914a

Rubio, D. M. G., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*, *27*(2), 94–104. https://doi.org/10.1093/swr/27.2.94

Shepard, L. A. (2020). Discussion of Part II: Should "Measurement" Have a Role in Teacher Learning about Classroom Assessment? In Susan M. Brookhart & J. H. Mcmillan (Eds.), *Classroom Assessment and Educational Measurement* (pp. 192–206). Taylor & Francis.

Tilden, V. P., Nelson, C. A., & May, B. A. (1990). Use of Qualitative Methods to Enhance Content Validity. *Nursing Research*, *39*(3), 172–175. https://doi.org/10.1097/00006199-199005000-00015

Willis, J., Adie, L., & Klenowski, V. (2013). Conceptualizing teachers' assessment literacies in an era of curriculum and assessment reform. *The Australian Educational Researcher*, *40*(2), 241–256. https://doi.org/10.1007/s13384-013-0089-9

Wilson, F. R., Pan, W., & Schumsky, D. A. (2012). Recalculation of the Critical Values for Lawshe's Content Validity Ratio. *Measurement and Evaluation in Counseling and Development*, *45*(3), 197–210. https://doi.org/10.1177/0748175612440286

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