ORIGINAL ARTICLE
Volume 14 Issue 3 2022
DOI: 10.21315/eimj2022.14.3.5
ARTICLE INFO
Received: 12-04-2021
Accepted: 11-04-2022
Online: 28-09-2022

Critical Thinking Transfer Practice Instrument: A Content Validity Calculation Steps Based on Expert Panel Evaluation

Noor Azura Mat Said[1], Siti Mariam Bujang[2], Nor Aishah Buang[3], Harlina Halizah Siraj@Ramli[2], Mohd Nasri Awang Besar[2]

[1]*Faculty of Medicine and Health Sciences, Universiti Sains Islam Malaysia, Negeri Sembilan, MALAYSIA*

[2]*Department of Medical Education, Faculty of Medicine, Universiti Kebangsaan Malaysia, Selangor, MALAYSIA*

[3]*Centre of STEM Enculturation, Faculty of Education, Universiti Kebangsaan Malaysia, Selangor, MALAYSIA*

To cite this article: Mat Said NA, Bujang SM, Buang NA, Siraj@Ramli HH, Awang Besar MN. Critical thinking transfer practice instrument: a content validity calculation steps based on expert panel evaluation. Education in Medicine Journal. 2022;14(3):61–74. https://doi.org/10.21315/eimj2022.14.3.5

To link to this article: https://doi.org/10.21315/eimj2022.14.3.5

ABSTRACT

Development of critical thinking instruments only focus on learning effectiveness, thus lacking the instrument measuring critical thinking transfer practice (CTTP) specific for medical undergraduates. This study aimed to develop an instrument that can measure CTTP amongst medical undergraduates in Malaysia with the primary focus on describing the steps for content validity calculation. A quantitative survey approach was applied through written questionnaire. Eight experts from public universities in Malaysia were selected through judgement sampling. The content validity calculation steps are presented in detail in form of content validity index (CVI) and modified Kappa coefficient (κ^*) score. The validity test involved 216 items, representing six sub-constructs. Item-level CVI (I-CVI), κ^* score and scale-level CVI (S-CVI) were calculated for evaluation criteria, namely, relevance, clarity and representativeness of each item. Results revealed that 213 out of 216 items were rated as appropriate (I-CVI $\geq$ 0.80) on all the evaluation criteria. From 213 items, 133 outstanding items were included in the instrument. After items deletion, the validity of the instrument improved. The study results may impact future development of critical thinking assessment in education. The instrument has the potential to be a CTTP psychometric measuring tool for medical undergraduates. The study will provides for others the knowledge of content validation process and procedures with emphasis on steps of content validity calculation. The study will also helps medical educators to better understand CTTP sub-constructs. These sub-constructs may help them to create better critical thinking pedagogies.

Keywords: *Critical thinking transfer, Medical undergraduates, Content validity, Expert evaluation*

CORRESPONDING AUTHOR

Siti Mariam Bujang, Department of Medical Education, Faculty of Medicine, Universiti Kebangsaan Malaysia, Jalan Yaacob Latif, 56000 Cheras, Kuala Lumpur

Email: ctmariam@ppukm.ukm.edu.my

INTRODUCTION

Critical thinking transfer is essential in students' learning especially when there is an educational transformation is required in learning the medical concepts. Nowadays, medical learning mainly focuses on analysis of the whole human body rather than subject-based learning approach (1). Educational transformation urges students to use knowledge and skills to analyse all medical concepts which helps them to comprehend information on human body. Apart from that, medical learning has changed from classroom learning to self-directed and blended learnings (2). This transformation urges medical undergradutes to transfer knowledge and skills that they have learnt from the classroom to new situations in order to survive in the real world.

Numerous studies have been conducted on critical thinking transfer (3–11). However, there were gaps in these studies. The first gap is that critical thinking transfer sub-constructs were not related to each other in reported theories (9–14). The second gap is that these sub-constructs were not specific to medical undergraduates' critical thinking learning. Besides, there is a lack of studies on instrument development that measures medical undergraduates' critical thinking transfer practice (CTTP). Previous studies mainly focused on instruments that measure learning effectiveness by measuring the student's critical thinking skills and dispositions (15–21). Thus, this study aimed to develop an instrument that measures medical undergraduates' CTTP. A separate CTTP instrument for medical undergraduates is needed because other established critical thinking instruments do not focus on critical thinking transfer that is important for medical undergraduates (15–29).

Hence, this study created a CTTP instrument for medical undergraduates. The researchers came out with prior construct, process of CTTP for medical undergraduates, and six prior sub-constructs that represented construct based on the literature and theories (3–14). The sub-constructs are: A – Different conceptualisations on critical thinking transfer concepts; B – Situations that allow critical thinking transfer; C – Understanding and examining deep structure of the problem; D – Rapid recognition of problem; E – Deliberate practice; and F – Learning strategies that promote critical thinking transfer. However, the prior construct and its sub-constructs were revised due to they were more relevant to non-medical context and connection between them have not been established. So, the final construct involved only medical undergraduates' CTTP. Also, the six sub-constructs were amended to: A – Conceptualising critical thinking transfer; B – Determining situations that allow critical thinking transfer; C – Understanding and examining deep structure of a problem; D – Recognising a problem rapidly; E – Practising critical thinking deliberately; and F – Determining strategies that promote critical thinking transfer.

The instrument was developed based on 10-phase development process. The phases were: (a) develop an initial thematic framework; (b) conduct a qualitative study; (c) fit the qualitative data into the initial thematic framework; (d) develop the operational definitions; (e) map the qualitative findings; (f) confirm the qualitative findings; (g) refine the qualitative data; (h) develop the CTTP constructs and sub-constructs; (i) develop the CTTP instrument; and (j) instrument testing.

In developing the CTTP instrument, content validity is one of the crucial validations that must be conducted. It is often done via review of the literature, interview of relevant populations and expert panel evaluation (30–38). It can also be achieved through three approaches, namely quantitative, qualitative, or mixed approaches (34, 39–45). In the quantitative approach, the panel may use consistency

and consensus strategies in evaluating the items (41–43). Metrics such as α coefficient, α consistency, κ statistics, content validation ratio, and Delphi are used to determine content validity scores for items in an instrument using mathematical calculations (41, 44, 46–48) where different metrics have different calculation methods. Thus, a content validation calculation is defined as a calculation method of a content validity score using a quantitative metric.

The quantitative metric used in the study was the content validity index (CVI) because it not only helps in making decisions on items (i.e., eliminating, modifying, or conserving them), but it is the most widely reported approach for content validity of an instrument (40–41, 44, 46, 49). The CVI calculation is simple, easy to understand and provides information about each item (44, 50). It also helps in evaluating the overall instrument through the calculation of scale-level CVI (S-CVI). The CVI is much easier to interpret compared to the content validation ratio. CVI determines items' status, whether the items are acceptable, needs revision or should be removed from the instrument.

Although content validation calculation is important, not many articles conduct this process (34, 39–40). Therefore, the purpose of this study was to report the steps of content validity calculation. The content validation of the CTTP instrument reported in the study is part of Phase 9 of the CTTP instrument development. There were five steps involved in the content validation of this instrument. The steps were described in the method section, sub-section content validation steps.

METHODOLOGY

CTTP Instrument

The instrument was developed for measuring the level of medical undergraduates' CTTP in a public university in Malaysia. The researchers developed the instrument, construct (medical undergraduates' CTTP) and sub-constructs found in the literature in Malay language. Initially, the instrument contained 216 items for all the sub-constructs (A : 40 items; B : 28 items; C : 51 items; D : 33 items; E : 30 items; F : 34 items).

Content Validation Steps

Content validation of the CTTP instrument reported in the study is part of Phase 9 of the CTTP instrument development. A summary of the stages of content validation of the CTTP instrument is reported in Phase 9. There were five stages involved in the content validation of this instrument.

Firstly, the aforementioned construct and six sub-constructs were conceptualised from the literature review. Secondly, the researchers consolidated the previously reported construct and six sub-constructs found in the literature. This consolidation helped the researchers to understand the connection between the construct and sub-constructs. In the third stage, the existence of the six sub-constructs was confirmed through a qualitative study. This confirmation helped to ensure that the construct and six sub-constructs were contextual to medical undergraduates learning process. An expert panel validated the six sub-constructs in the fourth stage, and the sub-constructs sufficiently represent the constructs (34, 39, 51–52). Thus, the expert panel evaluated the representativeness of items in the instrument, clarity of language and sentence structure of the items, and the relevancy of the items as suggested by the literature (49–53).

Finally, the items measuring the six sub-constructs were determined using CVI and modified κ coefficient (κ^*) score. κ^* score helped to determine the level of agreement for each item from outstanding (1.00), excellent (0.75–0.99), good (0.60–0.74), fair (0.40–0.59), or low (< 0.40) (44–54).

Detailed content validation calculation steps

Content validation of the CTTP instrument was based on the item-level content validation using item-level CVI (I-CVI) and κ* score. Validation of the whole CTTP instrument is based on S-CVI.

Item-level content validation calculation step

I-CVI was calculated based on the total number of experts who gave three or four point for each rating criterion divided by the total number of experts (N = 8) as in formula (i).

$$\text{I-CVI} = \frac{\text{Total number of experts that give 3 or 4 point}}{N} \quad \text{(i)}$$

Calculation of I-CVI is important to decide whether to use or remove an item. This item decision was based on the Lynn's (41) range: > 0.79 (appropriate); 0.70–0.79 (need revision); < 0.70 (eliminated). The items within the "appropriate" range were retained and included in the CTTP instrument.

Besides deciding on an item, I-CVI was also converted to a κ* score. This conversion helps in determining the level of agreement on the items. In deciding the κ* score for each item, the researchers have calculated the probability of chance agreement (Pc) using formula (ii) (28).

$$\text{Pc} = [N!/A!(N\text{-}A)!] \times 5N \quad \text{(ii)}$$

As the study by Polit et al. (44) did not clearly show how this formula was applied, this current study explains in detail how the Pc is calculated. N refers to the total number of experts, i.e., eight experts in this study. A is the total number of experts who gave 3 or 4 point for each evaluation criterion (relevancy [Rel], clarity [Cla] and representativeness [Rep]). $N!$ refers to the set of expert numbers. For example, if N = 3, so $N!$ (3!) = six because expert numbers 1, 2, and 3 can be arranged according to the following set of numbers: {123, 132, 213, 231, 312, 321}.

If the sum of N and A are the same, the number of experts who gave 3 or 4 point equals the total number of experts $(N–A)!$ = 0!. Score 0! is one because an empty set {} can only be arranged with one probability of an expert set. Because the total number of experts in this study is eight, therefore, $N!$ = 8! was calculated using the formula 8 × 7 × 6 × 5 × 4 × 3 × 2 × 1.

After determining the I-CVI and Pc scores, the κ* score was calculated for each item for all evaluation criteria using the following formula, κ* = (I-CVI–Pc)/(1–Pc). The κ* score was interpreted based on the range adapted from the literature (41, 44). Items that were within the low range are usually eliminated (45).

Scale-level content validation calculation step

Also, this study determined the S-CVI, one of the vital step in determining the instrument validity. The definitions on the S-CVI from the literature (46, 49) brought the idea on steps of the S-CVI calculations namely S-CVI universal agreement (S-CVI/UA) or S-CVI average (S-CVI/Ave). As the S-CVI/Ave is non-conservative and easy to interpret (46), thus, the researchers adapted the approach to calculate the S-CVI for the study instrument.

This study has described the content validity calculation aforementioned above by presenting the results based on the data analysis.

Methods Involved in the Calculation

Study design

This study applied a quantitative survey approach using a written questionnaire, as it is more economical and faster, and ensured that all of the questionnaires could be collected (55–56). The survey was given to an expert panel face-to-face because technique is useful, as any issues regarding the use of survey can be resolved immediately (57).

Sampling

Eight experts from various public universities in Malaysia were selected. Based on the previous studies' suggestions (34, 41, 45, 52), this study used judgement sampling where the sample was chosen based on the subjects' expertise. Judgement sampling is the most appropriate method for this study because the expert panel can be chosen by the researchers based on a specific criteria. There were two types of experts in this study, namely, professional experts and field experts (lay experts). The eight experts in this study consisted of five professional experts and three field experts.

Professional experts usually know the research field, have published on the issue, presented papers at national or international seminars, and have work experience and expertise to collaborate in the study (52–53). These experts help in identifying whether the instrument was well-developed for psychometric testing (52). In this study, the professional experts included experts who had expertise in critical thinking and instrument development. Other studies (52–53) supported this view, which stated that the selected professionals must be experts in the study area and in instrument development. At least one of the professional experts also need to be expert in linguistics area to evaluate the language used in each statement (58). Four professional experts were experts in critical thinking and instrument development, and one was a linguistics expert. The criteria used to select the professional experts were: (a) academic qualifications (i.e., having a doctorate in the field of expertise); (b) at least five years of experience in the area; (c) still working at the university; (d) actively teaching or publishing in the related area; and (e) the expert's consent to engage and commit to this study.

The other three experts appointed were field experts that appointed consisted of individuals who were most related to the concept to be measured and selected based on their representativeness of the population of this study. The use of the research subject as an expert was parallel with the other studies (34, 52). Also, practitioners at the studied institution can become a source that allows the development and revision of the CTTP instrument due to their work and involvement with the actual population of the study (52). The literature suggests that a field expert is a non-professional who has devoted time and energy to acquire expertise. A field expert might also be defined as "non-professionals who hold some basic knowledge about a subject." So, the field experts in this study may include dean, deputy dean, university programme head, university instructors, or students. The field expert group helps to evaluate item phrases, clarity of terms, and suggest essential and relevant items (52). The field experts in this study are the deputy head of a clinical department and a clinical lecturer in a public university in Malaysia. They have served in their roles for 11 and 15 years, respectively. They are also currently active in teaching medical students in the institution. Also, another field expert included, a medical education expert from another public university in Malaysia.

Procedures

Experts' invitations and documents for experts' evaluation

The researchers met with the experts to explain the purpose and procedures of the study. Also, the purpose of the meeting is to gain the experts' permission to be involved in the study. The panel of experts was given two weeks to evaluate the instrument. However, the actual assessment time took between one month to two months.

Each expert was provided with the following:

a. Information on the study, evaluation procedures, and criteria that these experts need to use.

b. An expert's demographic survey form.

c. An instrument evaluation form that contained detailed information about the objectives of the study, construct and sub-constructs to be measured, and operational definitions and measurement scales. In this evaluation form, a table containing the construct, sub-constructs, concepts to be measured, items and resources, and items for the evaluation criteria were attached.

Criteria for content validity of the items

The criteria for the evaluation of the items were Rel, Cla and Rep. Each criterion was measured on a 4-point Likert scale, as specified in Table 1. The Rel criterion refers to the relevancy between the items in this instrument to the construct and sub-constructs that the researchers want to measure. The Cla criterion refers to the clarity of the items in this instrument in terms of sentence structure, language and type of words used. For the Rep criterion, items were evaluated based on representativeness of the operational definitions.

Data Analysis

Content validity analysis for this study was based on I-CVI described in the item selection section. Also, the whole instrument's evaluation based on the S-CVI is described in the Determination of Instrument Content Validity section.

Item selection

CVI and κ^* score were used as they provide more information on the items. For example, with I-CVI, the study only considers the items that fall within the "appropriate" (I-CVI $\geq$ 0.80) range. The κ^* score helps by filtering the "appropriate" items based on their level of agreement. Only items that were within the outstanding rating were considered in the CTTP instrument.

Determination of instrument content validity

Next, CVI for the S-CVI was determined. This study adopted the following S-CVI definition, "the proportion of experts who score items as relevant or representative with either a 3 or 4" (49, p. 273). There are two S-CVI calculation methods, namely, SCVI/UA and S-CVI/Ave. Of the two ways, S-CVI/Ave is found to be less conservative and more easily interpreted (46). The latter approach was used to calculate the S-CVI value for the instrument. The value is calculated by the sum of the overall values of the CVI for the items (I-CVI) divided by the total number of items in the instrument.

Table 1: 4-point Likert scale in each evaluation criterion

Likert scale	Evaluation criteria		
	Rel	Cla	Rep
1	Not relevant	Not clear	Not representative
2	Relevant but need major revision	Clear but need major revision	Representative but need major revision
3	Relevant but need minor revision	Clear but need minor revision	Representative but need minor revision
4	Very relevant	Very clear	Very representative

RESULTS

The key findings in this study are presented in terms of analysis of the CVI and κ* score.

Number of Items in the First Draft of the CTTP Instrument

The first draft of the CTTP instrument comprised of 216 items with six amended sub-constructs: A – Conceptualising of critical thinking transfer; B – Determining situations that allow critical thinking transfer; C – Understanding and examining deep structure of a problem; D – Recognising a problem rapidly; E– Practicing critical thinking deliberately; and F – Determining strategies that promote critical thinking transfer. These sub-constructs are listed in Tables 2 and 3.

The list of items and their selection for the CTTP instrument was based on the I-CVI and κ* score. As there were initially 216 items in the CTTP instrument, this study only considered the "appropriate" items (I-CVI ≥ 0.80) and outstanding items (κ* = 1.00).

Item Status Based on I-CVI

Table 2 shows the evaluation classification and I-CVI score based on guidelines from previous studies (34, 49, 53). Based on these guidelines, there were 216 items for the Rel criteria, 213 items for Cla and 216 items for Rep within the range from 0.80 to 1.00 which are classified as appropriate (I-CVI ≥ 0.80) to be incorporated in the CTTP instrument. Only three items under the Cla criterion ranged from 0.70 to 0.79, which would need further revision, and no item was below 0.70. These three Cla items would have to be revised but based on the suggestion by the literature (49, 53), these items were unacceptable (I-CVI < 0.8) for the new instrument. Hence, these items were eliminated from the CTTP instrument, resulting in the CTTP instrument draft comprising of 213 items.

Item Status Based on the κ* Score

The items were then evaluated using the κ* score. Based on the results in Table 3, 177 items for Rel, 134 items for Cla and 176 items for Rep were rated as outstanding. The excellent items for Rel, Cla and Rep criteria were 39, 79, and 40 items, respectively. Three other items were rated as good (κ = 0.72) although their I-CVI was < 0.80. Thus, the authors recommended that those three items be eliminated. In the final part of the study, the researchers only considered the outstanding items (κ = 1.00) in all the evaluation criteria.

Based on the second evaluation using κ* score, only 133 items out of 216 items in Table 3 have been addressed by the content experts as outstanding items based on all the evaluation criteria (Rel, Cla and Rep). So, these 133 items were considered to be incorporated in the CTTP instrument.

S-CVI of the CTTP Instrument

Table 4 shows the calculation of content validity for the CTTP instrument through SCVI/Ave. S-VI/Ave score higher than 0.90 for an instrument is considered adequate content validity (50). Only 133 items with S-CVI/Ave = 1.00 were considered for inclusion in the CTTP instrument from the analysis. So, it is anticipated that the CTTP instrument has perfect S-CVI.

Table 2: The number of items based on the evaluation classification and I-CVI scores

Sub-constructs	Appropriate (> 0.79)			Need revision (0.70–0.79)			Eliminate (> 0.70)			Total number of items
	Rel	Cla	Rep	Rel	Cla	Rep	Rel	Cla	Rep	
A	40	40	40	0	0	0	0	0	0	40
B	28	27	28	0	1	0	0	0	0	28
C	51	50	51	0	1	0	0	0	0	51
D	33	33	33	0	0	0	0	0	0	33
E	30	29	30	0	1	0	0	0	0	30
F	34	34	34	0	0	0	0	0	0	34
Total	216	213	216	0	3	0	0	0	0	216

Note: Rel = Relevancy; Cal = Clarity; Rep = Representative

Table 3: The number of items based on the evaluation classification and $\kappa^{\star}$ score

Sub-constructs	Outstanding (1.00)			Excellent (0.75–0.99)			Good (0.60–0.74)			Fair (0.40–0.59)			Poor (< 0.40)		
	Rel	Cla	Rep	Rel	Cla	Rep	Rel	Cla	Rep	Rel	Cla	Rep	Rel	Cla	Rep
First evaluation															
A	38	36	38	2	4	2	0	0	0	0	0	0	0	0	0
B	20	19	21	8	8	7	0	1	0	0	0	0	0	0	0
C	49	14	48	2	36	3	0	1	0	0	0	0	0	0	0
D	18	18	18	15	15	15	0	0	0	0	0	0	0	0	0
E	21	22	21	9	7	9	0	1	0	0	0	0	0	0	0
F	31	25	30	3	9	4	0	0	0	0	0	0	0	0	0
Total	177	134	176	39	79	40	0	3	0	0	0	0	0	0	0
Second evaluation															
A	36	36	36	0	0	0	0	0	0	0	0	0	0	0	0
B	19	19	19	0	0	0	0	0	0	0	0	0	0	0	0
C	14	14	14	0	0	0	0	0	0	0	0	0	0	0	0
D	18	18	18	0	0	0	0	0	0	0	0	0	0	0	0
E	21	21	21	0	0	0	0	0	0	0	0	0	0	0	0
F	25	25	25	0	0	0	0	0	0	0	0	0	0	0	0
Total	133	133	133	0	0	0	0	0	0	0	0	0	0	0	0

Table 4: The calculation of S-CVI/Ave

	Rel			Cla			Rep		
	> 0.79	0.70–0.79	< 0.70	> 0.79	0.70–0.79	< 0.70	> 0.79	0.70–0.79	< 0.70
No. of item	133	0	0	133	0	0	133	0	0
I-CVI	133	0	0	133	0	0	133	0	0
S-CVI	1.00			1.00			1.00		

DISCUSSION

This study has developed an instrument to assess medical undergraduates' CTTP. In developing the instrument, the researchers have looked into established critical thinking instruments as the benchmark. In studies outside of Malaysia, the established critical thinking instruments focused on generally assessing the critical thinking ability of undergraduates. Some of the instruments focused on the critical thinking disposition constructs such as the California Critical Thinking Disposition Inventory (15), California Measure Mental Motivation (23), Critical Thinking Disposition Scale (17) and Yoon's Critical Thinking Disposition Instrument (18). Other established instruments focused on assessing the critical thinking skills of undergraduates, for example, Watson-Glaser Critical Thinking Appraisal (20, 24), Ennis-Weir Critical Thinking Essay Test (25), California Critical Thinking Skills Test (22), Halpern Critical Thinking Assessment (16), Collegiate Learning Assessment (59) and Cornell Critical Thinking Test (28). In Malaysia, the critical thinking instruments developed assessed on both dispositions and critical thinking proficiency of undergraduates such as Malaysian Critical Thinking (19) and Instrumen Kemahiran Pemikiran Kritikal (29). The established instruments have limited focus on the construct of critical thinking transfer. Thus, this study focused on the construct of CTTP.

In focusing on the CTTP, the researchers found a prior construct namely the process of CTTP of undergraduates based on several underlying theories (9–14). The researchers also came out with the six prior sub-constructs that represented the construct based on the literature (60–67). However, the construct and sub-constructs were primarily relevant for general learning context and the relationship between construct and sub-construct has never been established. Thus, the researchers conceptualised the construct and sub-constructs relevant to medical undergraduates learning context based on the combination of theories, past studies and a qualitative study. Based on researchers' conceptualisation, the study has given added value to the current knowledge regarding the relationship between the CTTP construct and sub-constructs. The prior construct was revised to make it more relevant to medical undergraduates' learning process. The final construct was the medical undergraduates' CTTP.

Based on the CTTP construct and sub-constructs, the study has contributed to advancing the current knowledge in critical thinking and instrument development area. Established instruments in the literature have limited focus on the CTTP construct. Thus, the study has developed an instrument that enables the assessment of CTTP construct based on the establishment of CTTP construct among learners. This instrument is significant for teachers and learners because it can be used to assess CTTP level among medical undergraduates. The assessment may help learners to improve their CTTP. Also, teachers may improvise their teaching and learning strategies in the area of critical thinking via results from this instrument.

Previous studies did not test the representativeness of the sub-constructs towards the construct. The testing is important to confirm content validity of the construct to be measured. Therefore, this study has contributed to the literature by testing the representativeness of the CTTP sub-constructs towards the construct through content validation process. Content validity of the CTTP instrument was developed based on three steps, namely, literature review, qualitative findings and evaluation of the expert panel as suggested by previous researchers (30–34).

This study presented several advantages of using content validity steps. In the expert panel evaluation, the content validity was helpful to assess the items in the instrument. The decision on whether to eliminate,

modify, or conserve the items was made using quantitative, qualitative and mixed approaches as suggested by previous researchers (33, 39–40, 42–43, 45, 68). The first strength is that this paper discussed the determination of the CTTP instrument's content validity. The validity is determined through expert panel evaluation using a quantitative approach. It is based on the experts' evaluation of an item – the experts' consistency estimates or their consensus on the rating scale for item relevance (41, 46–47). This study utilised the consensus approach instead of consistency approach because the latter has many disadvantages in item evaluation. Second, even there are numerous numbers of metrics used for the item evaluation in the consensus approach, such as α consistency, κ statistics, content validation ratio and Delphi (41, 44, 46, 48), the study used CVI because it is easy to compute, interpret and present in terms of item and instrument level analysis. The third strength was that this study has added value to previous studies by presenting the details of CVI calculations based on number of experts and converting CVI value to a κ^* score. Previous studies (34, 40, 50) did not show in detail on CVI calculation process. Fourth, this study has added value to the previous studies in which it applied three evaluation criteria (Rel, Cla, Rep) in evaluating the items. The criteria may give more data in evaluating the instrument validity as compared to other studies (34, 40) which included only two evaluation criteria. Fifth, this study has evaluated both I-CVI and S-CVI as compared to other studies. Other studies adapted either item level, or instrument level evaluations (34, 39–40, 42–45).

A possible drawback of this study is that the CVI does not consider the possibility of inflated values because of chance agreement (40, 46). Chance agreement refers to a concern in evaluating inter-rater agreement indices, especially when the choices are dichotomous, as is the case when 4-point ratings are collapsed into the two categories of relevant and not relevant (44). Hence the κ^* score provides additional CVI information as it does not adjust for chance agreement (46). The κ^* score helps in quantifying the level of agreement between experts (46).

CONCLUSION

CVI was used to determine the appropriate items to be included in the instrument. Then, the κ^* score was used to filter the items based on a defined level of performance. Items categorised as outstanding (I-CVI scored 1.00) were considered to be included in the instrument. The findings showed that the six sub-constructs could potentially be used to measure CTTP among medical undergraduates in Malaysia. In summary, the CTTP instrument with six sub-constructs and 133 items is ready to be performed to medical undergraduates. The study will benefit medical educators in understanding the CTTP sub-constructs in which these sub-constructs will help them create better critical thinking pedagogies. Also, the study provided specific content validation process and procedures that can be referred to and applied by other researchers. An added value offered by the study is that it emphasised the method of CVI calculation based on expert panel evaluation.

ACKNOWLEDGEMENTS

This study was supported by the Secretariat of Research and Innovation, Faculty of Medicine, Universiti Kebangsaan Malaysia with grant number FF-2017-436.

ETHICAL APPROVAL

This study was approved by Research Ethics Committee, Universiti Kebangsaan Malaysia [JEP-2017-642].

REFERENCES

1. Ismail R. Medical education in Malaysia: the evolving curriculum (part 2). J Clin Health Sc. 2017;2(1): 8–13. https://doi.org/10.24191/jchs.v2i1.5872

2. Ghee LK. The history of medicine and health in Malaysia. Selangor: Ghee LK; 2016.

3. Jerome C, Lee JAC, Ting SH. What students really need: instructional strategies that enhance Higher Order Thinking Skills (HOTS) among UNIMAS undergraduates. Int J Bus Soc. 2017;18(S4):661–8.

4. Kilbrink N, Bjurulf V, Baartman LKJ, de Bruijn E. Transfer of learning in Swedish technical vocational education: student experiences in the energy and industry programmes. J Vocat Educ Train. 2018;70(3):455–75. https://doi.org/10.1080/13636820.2018.1437064

5. Ludin SM. Does good critical thinking equal effective decision-making among critical care nurses? A cross-sectional survey. Intensive Crit Care Nurs. 2018;44:1–10. https://doi.org/10.1016/j.iccn.2017.06.002

6. Raaijmakers SF, Baars M, Paas F, van Merriënboer JJG, van Gog T. Training self-assessment and task-selection skills to foster self-regulated learning: do trained skills transfer across domains? Appl Cogn Psychol. 2018;32(2):270–7. https://doi.org/10.1002/acp.3392

7. van Peppen LM, Verkoeijen PPJL, Heijltjes AEG, Janssen EM, Koopmans D, van Gog T. Effects of self-explaining on learning and transfer of critical thinking skills. Front Educ. 2018;3(November):1–11. https://doi.org/10.3389/feduc.2018.00100

8. Wang C, Zhang X, Yao M. Enhancing Chinese college students' transfer of learning through service-learning. Stud High Educ. 2018;44(8):1316–31. https://doi.org/10.1080/03075079.2018.1435635

9. Halpern DF. Teaching critical thinking for transfer across domains. Am Psychol. 1998;53(4):449–55. https://doi.org/10.1037/0003-066X.53.4.449

10. Khalil MK, Elkhider IA. Applying learning theories and instructional design models for effective instruction. Adv Physiol Educ. 2016;40(2):147–56. https://doi.org/10.1152/advan.00138.2015

11. Perkins DN, Salomon G. Transfer of learning. In: Postlethwaite TN, Husen T, editors. International Encyclopedia of Education. 2nd ed. Oxford, UK: Pergamon Press; 1992. p. 1–13.

12. Kolb DA, Boyatzis RE, Mainemelis C. Experiential learning theory: previous research and new directions. Cleveland: Case Western Reserve University; 1999. p. 1–38.

13. Axeldrod R. Schema theory: an information processing model of perception and cognition. Am Polit Sci Rev. 1973;67:1248–66. https://doi.org/10.2307/1956546

14. Sweller J. Cognitive load during problem solving: effects on learning. Cogn Sci. 1988;12(2):257–85. https://doi.org/10.1207/s15516709cog1202_4

15. Facione PA, Facione NC. The California critical thinking dispositions inventory. California: California Academic Press; 1992.

16. Halpern DF. Halpern critical thinking assessment manual. Vienna, Austria: Schuhfried GmbH; 2010.

17. Irani T, Rudd R, Gallo M, Ricketts J, Friedel C, Rhoades E. Critical thinking instrumentation manual: UF/EMI UF/CT Quantitative Skills Test UF/CT Qualitative Skills Test. Florida: Rolfs Hall; 2007.

18. Shin H, Park CG, Kim H. Validation of Yoon's critical thinking disposition instrument. Asian Nurs Res. 2015;9(4):342–8. https://doi.org/10.1016/j.anr.2015.10.004

19. Abdul Rabu SN, Mohd. Mahdzir A, Aris B, Tasir Z. Assessing critical thinking skills of faculty education with MaCTIv4. Man India. 2017;97(17):105–19.

20. Watson G, Glaser E. Watson–Glaser critical thinking appraisal – UK edition practice test. London, England: Pearson Assessment; 2002. p. 1–15.

21. Yuan S, Liao H, Wang Y, Chou M, Medical CS. Development of a scale to measure the critical thinking disposition of medical care professionals. 2014;42(2):303–12. https://doi.org/10.2224/sbp.2014.42.2.303

22. Facione NC, Facione PA, Sanchez CA. Critical thinking disposition as a measure of competent clinical judgement: the development of the California Critical Thinking Inventory. J Nurs Educ. 1994;33(8):345–50. https://doi.org/10.3928/0148-4834-19941001-05

23. Giancarlo CA, Blohm SW, Urdan T. Assessing secondary students' disposition toward critical thinking: development of the California measure of mental motivation. Edu Pscho Measure. 2004; 64(2):347–64. https://doi.org/10.1177/0013164403258464

24. Gadzella BM, Baloğlu M. Psychometric properties of Watson-Glaser critical thinking appraisal for a sample of education majors. Psychol Rep. 2003;92(3 Pt 2):1249–54. https://doi.org/10.2466/pr0.2003.92.3c.1249

25. Ennis RH. Supplement to the test/manual entitled the Ennis–Weir Critical Thinking Essay Test. Urbana: Department of Educational Policy Studies, University of Illinois; 2005.

26. Facione PA. Critical thinking: a statement of expert consensus for purposes of educational assessment and instruction. Fullerton, CA: Insight Assessement; 1990.

27. Zahner D. Reliability and validity. New York: Council for Aid to Education; 2013.

28. Ennis R, Millman J, Tomko TN. Cornell critical thinking test level Z. Oregon: Critical Thinking Company; 2012.

29. Ariffin SR, Abdul Hamid NA. Profil kemahiran pemikiran kritikal antara pelajar aliran Sains dan bukan Sains. Malays Educ Dean's Counc J. 2009;3:20–7.

30. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976). 2000;25(24):3186–91. https://doi.org/10.1097/00007632-200012150-00014

31. Chang MC, Chen YC, Gau BS, Tzeng YF. Translation and validation of an instrument for measuring the suitability of health educational materials in Taiwan: suitability assessment of materials. J Nurs Res. 2014;22(1):61–8. https://doi.org/10.1097/jnr.0000000000000018

32. Squires A, Aiken LH, van de Heede K, Sermeus W, Bruyneel L, Lindqvist R, Brozstek T. A systematic survey instrument translation process for multi country, comparative health workforce studies. Int J Nurs Stud. 2012; 1(3):233–45.

33. Ho YY, Lim L. Targeting student learning needs: the development and preliminary validation of the learning needs questionnaire for a diverse university student population. High Educ Res Dev. 2021;40(7):1452–65. https://doi.org/10.1080/07294360.2020.1818062

34. Zamanzadeh V, Ghahramanian A, Rassouli M, Abbaszadeh A, Alavi-Majd H, Nikanfar A-R. Design and implementation content validity study: development of an instrument for measuring patient-centered communication. J Caring Sci. 2015;4(2):165–78. https://doi.org/10.15171/jcs.2015.017

35. Ayre C, Scally AJ. Critical values for Lawshe's content validity ratio: revisiting the original methods of calculation. Meas Eval Couns Dev. 2014;47(1):79–86. https://doi.org/10.1177/0748175613513808

36. Bolarinwa O. Principles and methods of validity and reliability testing of questionnaires used in social and health science researches. Niger Postgrad Med J. 2015;22(4):195. https://doi.org/10.4103/1117-1936.173959

37. McCoach DB, Gable RK, Madura JP. Instrument development in the affective domain. New York: Springer; 1986.

38. Chong P, Tham S, Kam JA. Measuring international students' satisfaction: the development of survey instrument. Global Conference on Business and Economics Research (GCBER) 2017, 14–15 August 2017, Universiti Putra Malaysia, Malaysia.

39. De Vellis RF, Thorpe CT. Scale development: theory and applications. Chapel Hill, US: SAGE Publications; 2017.

40. Hadzaman NAH, Takim R, Nawawi AH, Mohamad Yusuwan N. Content validity of governing in Building Information Modelling (BIM) implementation assessment instrument. IOP Conference Series: Earth and Environmental Science, Volume 140, 4th International Conference on Civil and Environmental Engineering for Sustainability (IConCEES 2017); 4–5 December 2017, Langkawi, Malaysia. https://doi.org/10.1088/1755-1315/140/1/012105

41. Lynn MR. Determination and quantification of content validity. Nurs Res. 1986;35(6):382–6. https://doi.org/10.1097/00006199-198611000-00017

42. Mohd Effendi Mohd Matore. Pembinaan instrumen kecerdasan menghadapi cabaran (IKBAR) pelajar politeknik menggunakan model Rasch. Pulau Pinang: Universiti Sains Malaysia; 2015.

43. Nunnally JC, Bernstein IH. Psychometric theory. 3rd ed. US: McGraw-Hill; 1994.

44. Polit DF, Beck T, Owen SV. Focus on research methods is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Res Nurs Heal. 2007;40(4):459–67. https://doi.org/10.1002/nur.20199

45. Wynd CA, Schmidt B, Schaefer MA. Two quantitative approaches for estimating content validity. West J Nurs Res. 2003;25(5):508–18. https://doi.org/10.1177/0193945903252998

46. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Res Nurs Health. 2006;31:341–54.

47. Waltz CF, Strickland OL, Land ER. Measurement in nursing and health research. 4th ed. New York: Springer Publishing Company; 2010.

48. Lawshe CH. A quantitative approach to content validity. Pers Psychol. 1975;28:563–75. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

49. Grant JS, Davis LL. Selection and use of content experts for instrument development. Res Nurs Health. 1997;20(3):269–74. https://doi.org/10.1002/(SICI)1098-240X(199706)20:3<269::AID-NUR9>3.0.CO;2-G

50. Polit DF, Beck CT. The content validity index: are you sure you know what's being reported? Critique and recommendations. Res Nurs Heal. 2006;29:489–97. https://doi.org/10.1002/nur.20147

51. Newman I, Lim J, Pineda F. Content validity using a mixed methods approach: its application and development through the use of a table of specifications methodology. J Mix Methods Res. 2013;7(3):243–60. https://doi.org/10.1177/1558689813476922

52. McGartland RD, Berg-Weger M, Tebb SS, Suzanne Lee E, Rauch S. Objectifying content validity: conducting a content validity study in social. Soc Work Res. 2003;27(2):94–104. https://doi.org/10.1093/swr/27.2.94

53. Davis LL. Instrument review: getting the most from a panel of experts. Appl Nurs Res. 1992;5(4):194–7. https://doi.org/10.1016/S0897-1897(05)80008-4

54. Domenic VC. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in Psychology. Psychol Assess. 1994;6(4):284–90. https://doi.org/10.1037/1040-3590.6.4.284

55. Creswell JD, Creswell JW. Research design: qualitative, quantitative, and mixed methods approaches. 4th ed. US: SAGE Publications; 2014.

56. Sekaran U, Bougie R. Research methods for bussiness a skill-bulding approach. 7th ed. West Sussex, UK: John Wiley & Sons Ltd.; 2016.

57. Ghazali HD, Hussin S. Metodologi penyelidikan dalam pendidikan amalan dan analisis kajian. 2th ed. Kuala Lumpur, Malaysia: Penerbit Universiti Malaya; 2016.

58. Sumintono B, Widhiarso W. Aplikasi pemodelan Rasch untuk penelitian ilmu-ilmu sosial. Cimahi, Indonesia: Trim Komunikata Publishing House; 2013.

59. Klein S, Benjamin R, Shavelson R, Bolus R. The collegiate learning assessment: fact and fantasies. Eval Rev. 2007;31(5):415–39. https://doi.org/10.1177/0193841X07303318

60. Anique B, Bruin D, Sibbald M, Monteiro S. The science of learning. In: Swanwick T, Forest K, O'Brien BC, editors. Medical education evidence, theory, and practice. 3rd ed. Oxford, UK: Wiley Blackwell; 2019.

61. Tiruneh DT, Gu X, De Cock M, Elen J. Systematic design of domain-specific instruction on near and far transfer of critical thinking skills. Int J Educ Res. 2018;87(October 2017):1–11. https://doi.org/10.1016/j.ijer.2017.10.005

62. Horn S, Veermans K. Critical thinking efficacy and transfer skills defend against 'fake news' at an international school in Finland. J Res Int Educ. 2019;18(1):23–41. https://doi.org/10.1177/1475240919830003

63. Mancinetti M, Guttormsenb S, Berendonkb C. Cognitive load in internal medicine: what every clinical teacher should know about cognitive load theory. Eur J Intern Med. 2018;1–5. https://doi.org/10.1016/j.ejim.2018.08.013

64. Leppink J. Cognitive load theory: practical implications and an important challenge. J Taibah Univ Med Sci. 2017;12(5):385–91. https://doi.org/10.1016/j.jtumed.2017.05.003

65. Robin J, Garzon L, Moscovitch M. Spontaneous memory retrieval varies based on familiarity with a spatial context. Cognition. 2019;190:81–92. https://doi.org/10.1016/j.cognition.2019.04.018

66. Sunde KE, Cutsforth-Gregory JK, Hunderfund ANL. The basic four approach to clinical neuroscience instruction: using cognitive load theory to enhance case-based learning. J Med Educ Curric Dev. 2018;5:1–6. https://doi.org/10.1177/2382120518815698

67. Kok EM, Geel K van, Merrienboer JJ van, Roben SG. What we do and do not know about teaching medical image interpretation. Front Psychol. 2017;8(309):1–8. https://doi.org/10.3389/fpsyg.2017.00309

68. Delgado-rico E, Carretero-dios H, Ruch W. Content validity evidences in test development: an applied perspectives. Int J Clin Heal Psychol. 2012;12(3):449–59.